NUCLEI SEGMENTATION AND COLOR NORMALIZATION OF HISTOLOGICAL IMAGES:

Rough-Fuzzy Circular Clustering to Deep Generative Modeling

A thesis submitted to Indian Statistical Institute
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy in Computer Science

by

Suman Mahapatra
Senior Research Fellow

Under the supervision of
Dr. Pradipta Maji, Professor

Machine Intelligence Unit
Indian Statistical Institute, Kolkata

December 2024

You cannot believe in God until you believe in yourself.

-Swami Vivekananda

Acknowledgements

Abstract

Histological image analysis deals with microscopic examination of tissue samples, which are stained with multiple histochemical reagents to highlight different cellular structures. The most important characteristic of histological images is the enormous amount of information, which makes computer-aided diagnosis more accurate than other imaging modalities. One of the important tasks in digital pathology is nuclei segmentation as it provides significant morphological information, which helps in therapeutic diagnosis and treatment of cancer. However, nuclei segmentation is a challenging task as nuclei structures can exhibit different morphologies, color, texture or can be occluded partially by other cellular components. Segmentation becomes more difficult when histological images suffer from inadmissible inter and intra-specimen disagreement in the appearance of stained tissue color due to inconsistency in staining routine, orientation of lens aperture, and so on. So, color normalization without hampering histological information is important for accurate nuclei segmentation. One of the main problems associated with stain color normalization is uncertainty due to incompleteness and vagueness in stain class definition, and overlapping characteristics of stains.

In this regard, the thesis introduces the concept of rough-fuzzy circular clustering for stain color normalization. It judiciously integrates the merits of both fuzzy and rough sets. While the theory of rough sets deals with uncertainty due to vagueness and incompleteness in stain class definition, fuzzy set handles overlapping nature of histochemical stains. The proposed rough-fuzzy circular clustering works on a weighted hue histogram, which considers both saturation and local neighborhood information of an image. A new dissimilarity measure is introduced to deal with the circular nature of the hue values. Being a generalization of existing algorithms, rough-fuzzy circular clustering facilitates accurate color normalization of histological images, while its integration with von Mises distribution provides better modeling of circular data and enables proper analysis of stained histological images.

In order to normalize the color without hampering the histological information of the image, a new deep generative model is introduced to capture the disentangled color appearance and stain bound information. To deal with the overlapping nature of histochemical stains, the proposed model assumes that the latent color appearance code is sampled from a mixture of truncated normal distributions. To segment cell nuclei from histological images, a new deep generative model is introduced, which considers an embedding space for handling information-asymmetry between histological image space and segmentation map. Judiciously integrating the concepts of optimal transport and measure theory, the model develops an invertible generator, which provides an efficient optimization framework with lower network complexity. Since segmentation and color normalization are two intertwined procedures, a novel simultaneous

segmentation and color normalization model is finally introduced, integrating the merits of spatial attention and truncated normal distribution. The latent color appearance information is assumed to be pairwise independent of nuclei segmentation map and other tissue-level details.

Contents

List of Figures

xiii

List of Tables

Chapter 1

Introduction

Medical imaging serves as one of the most notable advancements in modern diagnostic healthcare, which has revolutionized the way in which countless varieties of medical conditions are analyzed. This ever-advancing field develops different technologies to create visual representations of the anatomical and functional organization of a body for clinical analysis by providing non-invasive ways to diagnose internal structures. The technological revolution in the field of diagnostic healthcare started in 1895 with the discovery of X-ray imaging by German physicist Dr. Wilhelm Conrad Roentgen. Since then, this inventive field has experienced remarkable advancements which continue to evolve with the discovery of every new modality. Currently, a number of imaging modalities are available for medical analysis, such as, projection X-ray, echocardiogram, ultrasound, computed tomography (CT), positron emission tomography (PET), single photon emission computed tomography (SPECT), magnetic resonance imaging (MRI), functional magnetic resonance imaging (fMRI), pathological imaging, and so on.

Medical image analysis plays a very important role in the planning and execution of treatment strategies. For example, imaging modalities like PET scans and MRIs are used in the treatment of cancer to assess the grade of the disease and monitor therapeutic responses. This allows the medical practitioners to plan personalized treatment strategies that can be adjusted based on the effectiveness of the therapeutic intervention. Now-a-days, medical tests through different imaging modalities are generally prescribed in the follow-up of a disease, which has already been diagnosed and/or treated. Although clinical judgment plays a very important role in the treatment of different medical conditions, the use of medical imaging techniques is very crucial in detecting, confirming, assessing, and documenting courses of several diseases as well as planning subsequent treatment strategies. As per the World Health Organization (WHO), with upgraded healthcare policy and availability of modern medical instruments, a huge surge in medical imaging based treatment procedures is observed [1].

However, continuous advancement in the field of medical imaging has generated massive amount of data through several medical imaging devices and protocols, from which extracting meaningful information manually is almost nearly impossible. In this respect, computer-aided diagnosis (CAD) systems come into the picture for managing enormous volumes of medical data generated via modern healthcare systems. To process and analyze such massive data, the CAD systems utilize advanced machine learning and artificial intel-

ligence algorithms. The CAD systems can improve diagnostic efficiency and accuracy by identifying patterns, anomalies, and correlations that may be difficult for human specialists to recognize because of the complexity and volume of this massive amount of data. The CAD based systems make it possible to analyze medical images quickly and automatically, which eventually facilitates in early detection of diseases and improved patient outcomes. Hence, CAD systems are extremely useful for handling large amounts of medical data, streamlining diagnostic procedures, and improving the quality of healthcare as a whole, because of their scalability and accuracy.

1.1   Importance of Histological Image Analysis

Histological image analysis, which has its roots in clinical pathology, deals with the microscopic examination of tissue slides to provide unparalleled insights about the cellular structures, which makes it indispensable in contemporary medical diagnostics. The most important characteristic of histological images is the enormous amount of information, which makes CAD more accurate than radiological, cytological and other imaging modalities. Histological images are stained with multiple-contrasting histochemical reagents to highlight different cellular structures. Traditionally, pathologists have carried out the whole process, manually examining each slide meticulously to identify abnormalities within the tissue images. However, manual analysis is a time-consuming and labor-intensive task, which often leads to faulty diagnosis of diseases. Fig. 1.1 presents a step-by-step image acquisition procedure from histological tissue slides. The advent of digital pathology, where tissue slides are scanned using powerful microscopes to acquire high-resolution images, has paved the way for automated image analysis techniques. Histological image analysis possesses several advantages over other imaging modalities, which are listed as follows:

Figure 1.1: Image acquisition process from tissue slides.

- **Early Detection:** One of the main advantages of histological image analysis lies in the early detection of diseases. For example, in oncology, biopsy samples are examined histologically to detect the existence of cancer cells by analyzing abnormality in nuclear size and shape, allowing for an early diagnosis and course of treatment.

- **Accurate Diagnosis:** Another key aspect of histological image analysis is accurate diagnosis. Pathologists can recognize specific abnormalities in nuclear morphology

within the histological tissue images, which offer a thorough perspective of tissue structure and cellular morphology.

- **Treatment Strategy and Monitoring:** Treatment plans also heavily rely on histological image analysis. By delineating the boundaries of nuclei structures and tumors, and determining the degree of disease dissemination, a thorough examination of tissue samples aids in the guidance of surgical procedures.

- **Quantitative Analysis:** Another significant characteristic is quantitative analysis through cell counting and morphometric study based on segmented nuclei regions. The discovery of new therapeutic strategies is aided by quantitative analysis of histological data, which also helps in understanding disease mechanisms.

- **Medical Research and Drug Discovery:** In addition to its clinical uses, histological image analysis serves as an important tool in medical research and drug discovery by providing significant insights about the pathophysiology of diseases at a cellular level.

Therefore, histological image analysis plays a crucial role in medical imaging as it provides precise and in-depth understandings of tissue structure and pathology. Its significance extends to medical research, monitoring, early disease detection, precise diagnosis, and therapy planning. Histological image analysis has become a vital tool in contemporary healthcare because of the developments in artificial intelligence and machine learning, which enhances its already impressive capabilities.

1.2   Nuclei Segmentation

In histological image analysis, *nuclei segmentation* from histological images is an essential task for accurate diagnosis, prognosis, and detection of different diseases, especially cancer. This process entails locating and delineating the boundaries of cell nuclei within tissue samples. Accurate nuclei segmentation is a crucial step in histological image analysis as it helps the pathologists to quantify and examine cellular properties such as nuclear shape, size, density, and spatial distribution. These characteristics are essential markers of pathological states, aiding in the detection of malignant cells, comprehending the stage of the disease, and assessing the efficacy of treatment therapies.

Challenges in Nuclei Segmentation

Due to a number of inherent complexities, accurate segmentation of cell nuclei remains a challenging step in histological image analysis, despite substantial developments in digital pathology and computational approaches. The challenges associated with accurate segmentation of cell nuclei are listed as follows:

- **Variability in Nuclear Morphology:** Nuclei may exhibit different shapes, such as, spherical, elongated or oval, across different human organs. The size of cell nuclei can also vary significantly within a particle tissue sample. This diversity makes it very challenging to develop segmentation algorithms that can accurately detect and delineate potential nuclei structures from histological images.

- **Overlapping and Clustered Nuclei:** In closely packed tissue regions, nuclei structures may overlap, making it difficult for segmentation algorithms to identify the borders between individual nuclei. When two or more nuclei overlap, segmentation errors may occur as the overlapped nuclei structures are mistakenly recognized as a single entity. When neighboring nuclei come into contact with each other, nuclei structures form densely packed clusters. To prevent under-segmentation and guarantee accurate analysis, it is imperative to separate these clustered nuclei accurately.

- **Variation in Color Appearance:** There can be an observable disagreement in the color representation of stained tissue sections due to various factors, such as manual sectioning of the tissue samples, orientation of the lens aperture, inconsistency in staining routine, fading effect, types of microscopes used and so on. Segmentation becomes more difficult as this variability affects the contrast between nuclei structures, other cellular components and surrounding cytoplasm.

- **Presence of Artifacts and Noise:** Various artifacts, such as folds, bubbles, dust particles and debris, may be introduced during the tissue preparation process. Nuclear structures may be obscured or distorted by these aberrations, resulting in inaccurate segmentation of cell nuclei. Accurate nuclei detection can also be hampered in the presence of high noise levels within histological images, especially in low-contrast regions.

- **Heterogeneous Backgrounds:** The segmentation process becomes more challenging in the presence of heterogeneous backgrounds, such as, different cellular components which may be similar to cell nuclei in shape and size. Accurate segmentation requires algorithms to distinguish between nuclei and other cellular components.

- **Ground-Truth Variability:** Creating accurate ground-truth annotations for nuclei segmentation is a time-consuming, labor-intensive and highly subjective task, which leads to variability in the representations of reference data. This variability makes it very challenging to consistently assess and validate nuclei segmentation algorithms.

Addressing these aforementioned challenges requires continuous advancement in the discovery of novel computational techniques, and the development of robust and generalizable segmentation algorithms. Despite having these challenges, accurate nuclei segmentation is nevertheless very crucial for increasing diagnostic precision, improved patient outcomes, and advancing biomedical research.

1.3   Color Normalization

In digital pathology, *color normalization* of histological images is an important pre-processing step that aims to reduce the disagreement in the appearance of stained tissue color, caused by differences in staining routine, manual sectioning of tissue specimens and imaging conditions. Histological images are commonly stained with histochemical reagents, such as, hematoxylin and eosin (H&E), to highlight different cellular structures. However, inconsistencies in staining routine may lead to significant variations in stain color appearance. These variations can affect downstream image analysis tasks, such as, segmentation, feature

extraction and classification. Color normalization is an important step in histological image analysis, that reduces the inadmissible color variation corresponding to similar biological components within a single image or among the images within a biopsy set maintaining the structural and histological information contained in the images.

Challenges in Color Normalization

Despite its critical role in nuclei segmentation and subsequent image analysis tasks, achieving accurate results after color normalization is fraught with challenges. The challenges associated with color normalization of histological images are listed as follows:

- **Variability in Tissue Biology:** There is an inherent inconsistency in the structure and composition of biological tissues across different organs. This inconsistency may impact the way stains adhere to different biological components, leading to differences in color appearance, that are very challenging to standardize. Diseases can also alter tissue properties, which may affect the way reagents are absorbed, and eventually lead to differences in color representation.

- **Selecting Standard Reference:** Choosing a standard reference/template image or color space is another challenging task in color normalization. Due to the inherent inconsistency and pathological changes in biological tissues, differences in staining routine, it becomes very difficult to define a proper reference, which must represent a diagnostically meaningful color distribution.

- **Lack of Ground-truth Annotations:** Evaluating color normalization algorithms is another challenging task due to the lack of ground-truth annotations in histological image data sets. As the actual color representations of biological tissue regions are typically unknown, it becomes very hard to assess the accuracy of color normalization.

- **Subjective Assessment:** Assessment of color normalization often depends on visual evaluation by pathologists, which might result in inconsistencies as visual assessment can be influenced by subjective preferences and biases.

- **Adaptability and Robustness:** Another crucial aspect of color normalization is to ensure that the normalization algorithms generalize well across several types of staining protocols, different histological image data sets. Color normalization methods must be adaptable to new unseen data, and should be robust enough to handle extreme variations in staining, without introducing visual artifacts or losing important histological information.

To address the above challenges, active collaboration between pathologists and data scientists, and advancement in computational methods are very crucial.

1.4 Relevance of Rough-Fuzzy Clustering

One of the major problems in histological image analysis is uncertainty. Some of the sources of this uncertainty include incompleteness and vagueness in stain class definition,

as well as overlapping characteristics of multiple-contrasting histochemical stains. Hence, it is reasonable to use techniques or tools, that can handle and/or model these uncertainties efficiently in order to deal with stain color normalization problem of histological images. Both rough sets and fuzzy sets offer mathematical framework to capture and model uncertainties associated with real life data representations.

Rough set theory [147] provides an important mathematical tool to model ambiguity and uncertainty in data analysis. Unlike classical set theory, which requires crisp information, the theory of rough sets operates by synthesizing approximations of set concepts through upper and lower approximations. While lower approximation encompasses all the elements that definitely belong to that particular set, the concept of upper approximation includes those elements that possibly belong to the set. The boundary region, defined as the difference between upper and lower approximation, represents the uncertainty. On the other hand, fuzzy set theory [71] relies on the concept of a membership function on the domain of discourse. In order to represent an imprecise or overlapping concept, the fuzzy set theory assigns a degree of belongingness to each object. In granular universe, the concepts may be imprecise in nature as these concepts cannot be represented by crisp subsets. This leads to a direction, in which the principles of rough sets and fuzzy set theory can be integrated to develop a model of uncertainty, stronger than either of rough or fuzzy sets, under the notion of rough-fuzzy computing.

Rough-fuzzy clustering is an efficient hybrid approach that judiciously integrates the notions of rough sets and fuzzy sets to address the limitations of traditional clustering methods. While the membership functions of fuzzy sets enable efficient handling of overlapping classes, the notion of lower approximation and boundary region of rough sets deals with uncertainty, vagueness, and incompleteness in class definitions. Since the rough-fuzzy techniques can offer stronger paradigm for addressing uncertainty than either of the rough sets or fuzzy sets, it has more potential in the application domains of medical imaging. In this respect, the concepts of rough sets and fuzzy set theory have been integrated judiciously in this thesis to model and handle uncertainties in stain color normalization task.

1.5   Relevance of Deep Generative Modeling

Deep generative models have emerged as key innovation in the field of artificial intelligence and machine learning due to their extensive skills in generating, modeling, and comprehending complex data distributions. Deep generative models, namely, generative adversarial networks (GANs), variational autoencoders (VAEs), normalizing flows, and autoregressive models, allow for the generation of highly realistic synthetic data that closely resembles real-world data representation. Additionally, deep generative models play a pivotal role in data augmentation, increasing the number and diversity of training data sets to boost performance and robustness of machine learning algorithms. Despite its phenomenal success in every research domains, there are certain aspects which hinder the construction as well as overall performance of deep generative models. Some of these concerns are discussed below.

- **Selection of Architecture:** Determination of architecture for a deep generative model is an empirical issue. There does not exist any guiding principle which helps

in deciding the proper architecture for a model. The number of neurons in a particular layer or the number of layers in a deep model is generally heuristically determined based on evolutionary algorithms or existing pre-trained networks. However, the network architecture of a deep generative model, determined with such an approach, does not consider the diversities present in the nature of problems as well as underlying complexities associated with the given input data.

- **Overfitting:** An undesirable behavior of deep generative model is that it often attempts to fit the entire training data, and in that process, ends up memorizing the patterns, noise and random fluctuations present in the data. As a result, the models exhibit strong performance on training data, but fails miserably on unobserved data. This hinders the generalizability of the models as the models learn specific low-level features that may not be relevant for other domains or tasks. Overfitting occurs due to a number of factors, which are listed as follows:

  - irrelevant or redundant samples present in the training data,
  - the developed model is highly complex and is capable enough of learning the noise present in training data,
  - the model is trained on same set of samples over a long period of time: no proper early stopping criteria,
  - most importantly, the number of training samples is too small to reflect the underlying data distributions accurately.

  The last issue is very common in different types of medical image analysis tasks. These factors degrade the overall performance of deep generative models.

- **Computationally Expensive:** As deep generative models involve complex mathematical operations, multiple layers of neuron and a huge number of hyper-parameters, training and deploying them can be time-consuming and computationally very expensive compared to classical machine learning algorithms. A large amount of computational resources, such as, powerful central processing units, storage, high-performance graphics processing units, are essential to train most of the deep generative models. The amount of computational resources, required to efficiently train a deep generative models, generally depends on the depth and complexity of the model.

- **Lack of Interpretability:** Another challenging issue of deep generative models is the lack of interpretability, in which the outputs provided by the models are difficult to be interpreted. For this reason, deep generative models or in a broader sense, deep learning models are often considered as black boxes. It is still an empirical issue to analyze how it is determined that which feature is to be extracted at which layer for modeling the underlying distribution of a given input data. This issue can pose problems in application domains, such as medical image analysis, that require transparency and accountability.

- **Security:** Another challenging aspect of deep generative models is security. Deep learning models are vulnerable to small perturbations, that are generally not perceived by human eyes. Small additive noise added to an input image can fool a

already trained model and lead to misclassification by the model. Hence, the study about these adversarial attacks and robustness of the developed models against these attacks have become an important aspect of modern deep learning research.

Despite having these aforementioned issues, deep generative models have achieved significant advancement in numerous applications with exceptional performance. This is because of the fact that deep models perceive the environment as a nested hierarchy of concepts and representations, where each concept is defined in relation with much simpler concepts. This allows the deep models to attain remarkable power and flexibility. Deep generative modeling has a major influence on medical imaging research because of its ability to generate realistic synthetic images, that can be further utilized for training, facilitating data augmentation.

1.6   Scope and Organization of the Thesis

In this regard, this thesis addresses the problem of nuclei segmentation from histological images, in the presence of inadmissible inter and intra-specimen disagreement in tissue color appearance of stained histological images.
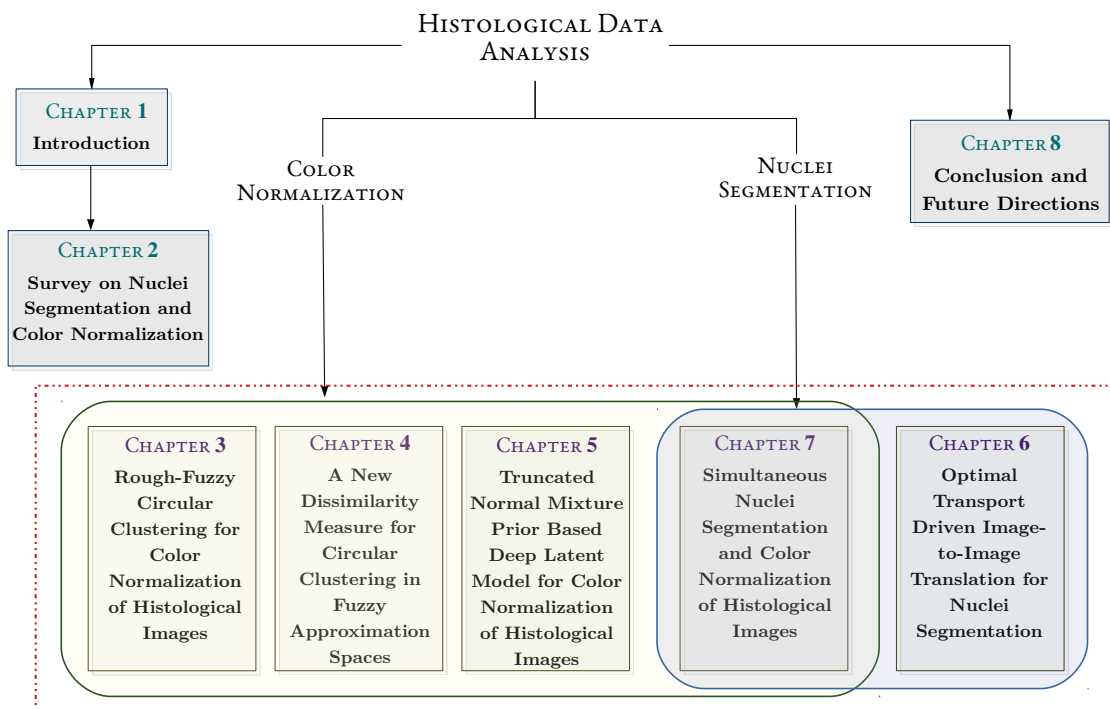
Figure 1.2: Outline of the thesis.

The schematic outline of the thesis is presented in Fig. 1.2. The thesis consists of eight chapters. Chapter 1 provides an introduction to medical imaging. The importance

of histological image analysis, significance of nuclei segmentation and color normalization, along with the relevance of both rough-fuzzy clustering and deep generative modeling in stain color normalization and nuclei segmentation of histological images, are also briefly described in Chapter 1. Chapter 2 covers a concise survey on state-of-the-art methods for nuclei segmentation and stain color normalization of histological images.

During the image acquisition process, tissue samples are stained with multiple contrasting histochemical reagents for highlighting different tissue structures and cellular features. Color disagreement among histological images may affect the performance of computer-aided histological image analysis. In this respect, Chapter 3 introduces a new circular clustering algorithm, termed as rough-fuzzy circular clustering ($RFCC_{cosine}$), for stain color normalization of histological images. The $RFCC_{cosine}$ algorithm incorporates the merits of both rough sets and fuzzy sets. The proposed algorithm assumes that each stain class is represented by a centroid, a crisp lower approximation, and a fuzzy boundary. While the membership function of fuzzy set enables efficient handling of overlapping stain classes, the concept of lower approximation and boundary region of rough sets deals with uncertainty, vagueness, and incompleteness in stain class definition. To deal with the circular nature of hue values, the cosine distance is considered. The proposed circular clustering algorithm also incorporates saturation-weighted hue histogram, which considers both saturation and hue information of a given histological image, to reduce the impact of less-saturated achromatic pixels on stain color estimation. Some new quantitative measures are also proposed to evaluate both within-image and between-image color constancy after normalization. The efficacy of the proposed method, along with a comparison with other state-of-the-art color normalization methods, is demonstrated on publicly available hematoxylin and eosin (H&E) stained benchmark histological images.

The cosine distance used in Chapter 3 is defined based on the cosine function. However, the cosine function fails to model concentration of circular hue values near the mean of each stain class. The area under the curve of cosine function can be varied, without affecting its periodicity, by multiplying any positive parameter. Based on this concept, a new dissimilarity measure is proposed in Chapter 4 to deal with the circular nature of hue values and to capture underlying data distribution properly. Based on the new dissimilarity measure, Chapter 4 introduces a new rough-fuzzy circular clustering algorithm, termed as $RFCC_{vM}$ for color normalization of histological images. In the proposed rough-fuzzy circular clustering, the estimated fuzzy membership function follows von Mises distribution, which is a variant of the normal distribution in circular domain. Integrating judiciously the merits of both saturation-weighted hue histogram and local neighborhood information of hue image, the $RFCC_{vM}$ algorithm works on a weighted hue histogram in HSI color space. The performance of the proposed method, along with a comparison with other state-of-the-art stain separation and color normalization methods, is demonstrated on publicly available standard data sets consisting of H&E stained histological images.

The rough-fuzzy circular clustering based stain color normalization methods, proposed in Chapter 3 and Chapter 4, do not explore the correlation between color appearance and stain bound information, corresponding to each histological image. Hence, it can not verified that the loss or modification in one information does not affect the other one, which may eventually lead to loss of histological information after color normalization. In order to normalize the color without hampering the histological information of the image, Chapter 5 introduces a new deep generative model, termed as $\underline{T}$runcated normal $\underline{M}$ixture prior

based Latent model (TredMiL) to reduce the color variation present among the histological images. The proposed model assumes that the latent color appearance information, extracted through a color appearance encoder, and stain bound information, extracted via stain density encoder, are independent of each other. In order to capture the disentangled color appearance and stain bound information, a generative module as well as a reconstructive module are considered in the proposed model. The discriminator is modeled to discriminate between not only the image samples, but also the joint distributions corresponding to image samples, color appearance information and stain bound information, which are sampled individually from different source distributions. To deal with the overlapping nature of histochemical reagents, the proposed model assumes that the latent color appearance code is sampled from a mixture model. As the outer tails of a mixture model do not contribute adequately in handling overlapping information, rather are prone to outliers, a mixture of truncated normal distributions is used to deal with the overlapping nature of histochemical stains. The performance of the proposed model, along with a comparison with state-of-the-art stain separation and color normalization approaches, is demonstrated on several publicly available data sets containing H&E stained histological images.

Chapter 5 demonstrates that color normalization has a huge impact on subsequent nuclei segmentation task. Mitigating the disagreement in the appearance of stained tissue color improves quality of region-of-interest (ROI) segmentation from histological images. After addressing the color normalization problem, the segmentation of regions-of-interest from histological images is a very crucial task for computer-aided histological image analysis technique. To devise a nuclei segmentation algorithm, generalizable to segmentation problems with different types of target map representations, image-to-image translation paradigm can be considered as they are invariant to target domain image representations. But, image-to-image translation models fail miserably when the information content between two image domains are asymmetric in nature. In this regard, Chapter 6 introduces a new deep generative model, termed as Optimal Transport Driven Asymmetric Image-to-Image Translation approach (Ostrich), for segmenting nuclei structures from histological images. The proposed Ostrich model considers an embedding space for handling information-disparity between information-rich histological image space and information-poor segmentation map domain. Integrating judiciously the concepts of optimal transport and measure theory, the model develops an invertible generator, which provides an efficient optimization framework with lower network complexity. The concept of invertible generator automatically eliminates the need of any implicit cycle-consistency loss term. The discriminator is modeled to distinguish between real and generated/fake pairs of segmentation map, and embedding information. The proposed model introduces a spatially-constrained squeeze operation within the framework of invertible generator to maintain spatial continuity within the image patches. An important finding is that the model provides a better trade-off between network complexity and model performance compared to other image-to-image translation models having complex network architectures. The effectiveness of the proposed deep generative model, along with a comparison with other state-of-the-art nuclei segmentation models, is demonstrated on publicly available histological image data sets.

Stain color normalization is one of the primary and necessary steps in histological image analysis, which enables better nuclei segmentation from histological images. On the other

hand, accurate nuclei segmentation makes color normalization rather trivial. Hence, these two intertwined procedures are integrated so that they take advantage from each other, simultaneously yielding better nuclei segmentation as well as better color normalization. In this regard, Chapter 7 introduces a simultaneous nuclei segmentation and color normalization model, which integrates the merits of spatial attention and truncated normal distribution. The concept of spatial attention is incorporated to segment nuclei regions from histological images. Similar to Chapter 5, the proposed model also assumes that the latent color appearance code is sampled from a mixture of truncated normal distributions. The latent color appearance information is assumed to be pairwise independent of nuclei segmentation map and other tissue-level details. The efficacy of the proposed model, along with a comparison with existing color normalization and nuclei segmentation approaches, is demonstrated on publicly available data sets containing H&E stained histological images.

Finally, the thesis is concluded in Chapter 8 with discussions on future scopes and improvements of the proposed research works.

Chapter 2

Survey on Nuclei Segmentation and Color Normalization

This chapter presents a succinct literature survey on nuclei segmentation and color normalization of histological images.

One of the important tasks in digital pathology is nuclei segmentation as it provides significant morphological information, which helps in therapeutic diagnosis and prognosis of cancer. However, nuclei segmentation is a challenging task as nuclei structures can exhibit different shape, size, color, texture or can be occluded partially by other cellular components. Segmentation becomes more difficult when histological images suffer from inadmissible inter and intra-specimen disagreement in the appearance of stained tissue color due to manual sectioning of the tissue specimens, inconsistency in staining routine, orientation of lens aperture, and so on. So, color normalization without hampering histological information is important for accurate nuclei segmentation. Next section presents a concise survey on stain color normalization approaches, followed by a survey on state-of-the-art nuclei segmentation methods.

2.1 Stain Color Normalization Approaches

Based on the strategy used to address the color disagreement problem in the appearance of stained tissue color, state-of-the-art stain color normalization methods can be broadly categorized into the following categories:

2.1.1 Histogram transformation based approaches

To standardize the color appearance of histological images, histogram transformation based approaches transform the input image color space to a reference RGB color space with a known color distribution. A large collection of non-normalized images, representative of the staining protocol, are used to derive the reference color space. In [38], the background color of the target image was transferred to that of the source image by first converting the RGB image to the $L\alpha\beta$ color space. Effective color correction is then achieved by converting the

image back to the RGB color space from the L$\alpha\beta$ color space. A method proposed in [146] normalizes the global illumination and saturation of images into HSV (hue-saturation-value) color space using precise adjustments to the saturation and value channels. Another method proposed in [144] normalizes the images using color transformation to a template image and linear projection in L$\alpha\beta$ color space. Stain normalization was achieved in [] by utilizing sparse autoencoders for standardization of color distribution. These sparse autoencoders generate features that are then used to cluster the image pixels. The pixels belonging to the these clusters are transformed to template image using a specific histogram projection approach.

A fuzzy-based color normalization method was introduced in [11], which can be considered as an improvement over the Reinhard method [38]. This method deals with the limitations of conventional approach by utilizing fuzzy logic, which controls contrast enhancement in the L space and color coefficients in the $\alpha\beta$ space. Another method proposed in [119] integrates fuzzy logic, enhanced fuzzy illuminant, and modified spectral normalization for resolving discrepancies in biopsy imaging and color disagreement. As histogram transformation based methods do not consider staining properties of the histopathology images, they generate undesired artifacts after color normalization [121].

2.1.2 Separated transformation based approaches

Based on the fact that whole slide images are stained using separate staining agents, some approaches based on multiple transformations were proposed in [15], [35], [25]. These approaches use matrix decomposition and color deconvolution technique to classify the pixels as distinct stains [7]. In order to separate light-absorbing stains, this method applies a calibration-based technique which utilizes the principles of Beer-Lambert law of colorimetry [7], [123]. A non-linear mapping based color deconvolution method was proposed in [15]. After removing the pixels that are part of the same stain, the stain distributions of the retrieved pixels are examined to estimate the staining vector. Color of different types of pixels is transferred to the template image using staining vector extracted via the color deconvolution method.

A hue saturation density based approach was proposed in [24] for standardizing the color representation of whole slide images. This method transforms the RGB color space into chromatic and density components. The chromatic and density distributions corresponding to each component in the hue saturation density color space are matched with their respective distributions by alignment. The approaches based on separated transformations normalize the image pixels by applying distinct transformations, which might introduce color discontinuities within the normalized images. This leads to incorrect structural alterations in stained tissue slides. To normalize the images through multiple operations, additional computation is also needed.

2.1.3 Unified transformation based approaches

The main disadvantage of the aforementioned methods is the discontinuity in color representation and the high computational time. In this respect, a large number of methods were proposed to establish a unified transformation for the pixels in histopathological images. For stain separation of histological images, color deconvolution method [7] was considered

as an effective technique in some methods [94], [142], [143]. Several color normalization approaches have been utilized, where color deconvolution and its inverse operation are used to identify the transformation between the source and template images [84], [86].

In [84], authors proposed a fringing approach that transforms RGB images into optical density space. In this approach, a plane was created based on the largest singular values and then the corresponding angle is estimated. In [151], the authors proposed a modified color deconvolution method through optimization of whole slide images. Similar to color deconvolution, to estimate optimized hematoxylin decomposition on localized images, optimization variable is used in this method. In [133], the authors used stain color normalization matrix over color deconvolution matrix as a model variable and applied non-negative matrix factorization to solve the underlying problems. To assign each pixel to a particular stain class, sparsity constraint was incorporated in non-negative matrix factorization framework [22], [77], [63]. A geometry-inspired chemical-invariant and tissue-invariant stain color normalization method was proposed in [9]. Utilizing underlying geometry-inspired color vector space, the variations in illumination, stain color and stain amount were rectified within a unified framework. A stain color adaptive normalization approach was proposed in [88], that uses clustering and segmentation for detecting histological structures without hampering the dominant colors.

2.1.4 Generative adversarial network based approaches

Recently, generative adversarial networks (GAN) have emerged as an important technique, which have been used in different ways to solve the stain color normalization problem. In [6], the authors combined a GAN with a task-specific network to perform stain transformation between images. In another GAN based method proposed in [47], the authors used grayscale images as input and utilized a classifier to improve performance of stain transformation. Both of these approaches rely on labeled data and are not generic. Based on the InfoGAN [130] model, an unsupervised generative model was developed in [41], that only requires labels corresponding to target stain style to perform stain color transformation. Based on the concept of CycleGAN [64] model, another method was proposed in [90] to address the stain color normalization problem. In [50], a pre-trained convolutional neural network, along with a GAN was utilized to normalization the color variation of whole slide images. A stain transformation method was proposed in [49] which uses a directional statistics-based color similarity index in conjunction with GAN. In [48], the authors proposed a robust and computationally efficient stain color normalization model. To retain semantic characteristics, a method was proposed in [39] that combines CycleGAN with a pre-trained segmentation network. A feature map loss is used in this method to discriminate between real and generated images. In [117], the authors investigated the utilization of CycleGAN with pathology-constrained representation network to enhance the model's capacity to learn pathological features.

2.2 Nuclei Segmentation Approaches

Due to the powerful feature extraction capability, deep learning methods have gained a lot of attention and have been applied in different types of medical image analysis tasks.

Hence, nuclei segmentation methods can be broadly categorized into two groups, which are discussed next.

2.2.1 Non-deep approaches

Though, recent years have witnessed a huge surge in the application of deep learning methods in many medical image segmentation, but non-deep methods have their own advantages. Classical methods are generally faster and they do not need huge number of sample for training the models. Some of the earlier works in this field have relied on energy-based methods, particularly watershed algorithm. In [92], the authors have used fast radial symmetry transform to extract the markers and utilized a sequence of morphological operations to extract energy landscape. A variant of the watershed algorithm was proposed in [112] that solely depends on energy landscape, and uses a combination of active contour and shape prior to extract nuclear regions. Utilizing the geometry of nuclear structures, a method was proposed in [62], which computes the concavity of nuclei clusters to perform the nuclei segmentation task. A cell/nuclei segmentation approach, based on the combination of double thresholding method and morphological operations, has been proposed in [29].

To address pathological image segmentation problem, some shallow learning based models have also been proposed in [61], [60], which perform convolution operation directly on the input histological image and try to extract pixel-wise labels through their abstract feature representations. The main issue with these approaches is that they do not consider spatial dependencies between pixels and hence capture inadequate information.

2.2.2 Deep learning based approaches

Segmentation of cell nuclei from histological images play a very important role in computer-aided medical diagnosis [37]. Several methods of quantifying cells in microscopy images, like cell counting and shape analysis, highly rely on accurate segmentation. Nuclei segmentation plays a pivotal role in detecting the gradation of cancers [28]. The information regarding segmented nuclei regions is considered as an important indicator of cancer prognosis [5]. Nuclei as well as cell segmentation is a very challenging task due to the the variability in shapes, sizes, staining, texture etc. Deep learning based approaches have achieved tremendous success in different types of medical image segmentation tasks. Deep learning based models can learn discriminative features, which makes them more effective in medical image segmentation than the classical methods. The deep learning based nuclei segmentation approaches can be broadly categorized into two groups based on the type of segmentation.

2.2.2.1 Semantic Segmentation

In semantic segmentation, each pixel p of an image is assigned to particular class c where, $c \in \{c_1, c_2 \cdots c_n\}$ for n number of classes [91], [135]. Deep learning based semantic segmentation methods can be categorized into different groups, some of which are enlisted as follows:

1. **Superpixel based methods:** High resolution histological images require large amount of memory and huge computation resources, such as high-performance central processing units (CPUs) or high-end graphics processing units (GPUs). In [120],

the authors have proposed a superpixel based deep learning model for learning localized features efficiently. This approach uses simple linear interactive clustering (SLIC) algorithm, proposed in [111] to extract superpixel information. A method similar to [120] was proposed in [21] to segment cell based on extracted superpixel patches. An resolution adaptive deep hierarchical learning (RADHicaL) approach was proposed in [12] for segmentation of nuclei regions. In [14], the authors trained M AlexNets with different resolutions and extracted the maximum of probabilities as the final output.

2. **Methods to deal with touching and overlapping nuclei:** Segmentation of nuclei structures in the presence of touching or overlapped nuclei in dense regions of histological images is a challenging task. In [97], the authors have addressed this issue for segmenting nuclei structures from different human organs. This method predicts ternary probability maps for background, nuclei structures and nuclear boundaries to perform accurate segmentation of cell nuclei instead of predicting touching or overlapped nuclei as a single entity. A cell and nuclei segmentation approach based on deep content-aware network (DCAN) was proposed in [46]. While the weights on the downsampling path were initialized using a DeepLab model [72] trained on the PASCAL VOC 2012 data set [80], the weights on the upsampling path were randomly initialized using a standard normal distribution. For improving nuclei segmentation in ganglion H&E stained histopathological images, a boundary-enhanced segmentation network (BESNet) was proposed in [52]. A boundary-enhanced loss term was introduced in this method to penalize the outputs, that deviates from ground-truth annotations. To address the problem of touching or overlapping nuclei in a dense region within a histological image, a regression based nuclei segmentation method was proposed in [107]. This regression based method generates distance maps to deal with the issue of overlapping nuclei, which can considered as an alternative method for predicting nuclei structures and the contours [31]. The regression based deep networks for generating distance maps are fully connected networks (FCNs) with ImageNet [57], [101], U-Net [40], and pre-trained Mask-R-CNN with ResNet 101 backbone [67], trained on common objects in context (COCO) data set [126], where to fine-tune the network weights two data sets, [97] and [106] are used. To segment accurate nuclei structures, morphology dynamics [82] is used on the estimated distance maps. Sparse reconstruction technique and deep learning methodology was used in [137] to segment nuclei structures from H&E stained breast cancer histopathology images. Another method, proposed in [138], is an fully connected network, that contains encoderâĂŞdecoder pathways akin to U-Net [40]. This approach segments borders and nuclei at the same time. Random patches of sparse non-negative matrix factorization based stain normalized H&E images serve as the inputs to the model [23]. In [17], the authors proposed a deep learning model that employs two U-Net architectures to segment the cells and generate a regression-based distance map in order to improve the separation of cells in segmented masks. This method performs stain color normalization on H&E stained images using [84] and then trains one of the U-Net models for segmentation. The distance map was created by training another U-Net. In order to segment cells from a H&E stained tissue images, a method was proposed in [65]. For this purpose, the method trains a U-Net

architecture with overlapping image patches.

3. **Cascaded deep learning approaches:** Wang et al. suggested a cascaded deep learning strategy to segment nuclei in brain tissue histology images [110]. In order to improve the segmentation quality, this method trained two CNNs. The probability maps estimated by the first neural network were fed to the second one, which used the information from the original image patches and the probability maps to learn high-level contextual information.

4. **Evaluation based studies:** In [95], the authors studied the impact of spatial information using three deep architectures. The aim of the study was to segment cellular and extracellular objects from histopathological tissue images using CNN, sparse autoencoder and deep belief network (DBN). An experimental study of nuclei segmentation from histological images was performed in [106] using three distinct deep networks, namely, a a shallow network called PangNet [26], a fully convolutional neural network [57], and DeconvNet [51]. Morphological dynamics [82] was used to post-process the estimated posterior probability maps.

5. **Recurrent neural network based segmentation:** In [122], a deep learning architecture called Her2Net was introduced for the purpose of classifying and segmenting cells and cell membranes on breast cancer images that were stained with human epidermal growth factor receptor-2 (HER2). Her2Net was developed an encoderâĂŞdecoder architecture that is comprised of sixteen convolutional layers, two max-pooling layers, two spatial pyramid pooling layers, and a trapezoidal long short-term memory (TLSTM). This network outperformed SegNet [127], Bayesian SegNet [13], and U-Net model [40] with respect to the Her2 data set.

6. **Adversarial neural network based approach:** In [3], an adversarial neural network based method was presented for segmenting cells from images. Generative adversarial networks (GANs) [54] serves as the basis for this method. Rather than creating images from a random noise vector, the method proposed in [3] estimates the segmentation map of cells. In [85], the authors suggested a method for training neural networks with millions of parameters using less annotated data.

2.2.2.2 Instance Segmentation

Instance segmentation can be defined as the process of identifying each object, followed by segmenting each detected object [135]. Instance segmentation methods based on deep architectures can be broadly categorized into two groups, which are listed as follows:

1. **Bounding box based instance segmentation:** A composite deep learning technique was proposed in [141] for the automated segmentation of cervical nuclei, which is followed by post-processing with the local fully connected conditional random field (LFCCRF). This method offers detection-based instance segmentation strategy through the application of mask R-CNN model [66]. Mask R-CNN includes more low resolution semantic features, which result in some boundary loss on the predicted segmentation masks. So, LFCCRF was used to refine the segmented masks.

2. **Centroid based instance segmentation:** Though deep learning has outperformed its classical counterpart in medical imaging research, the process of annotating data is subjective, laborious, prone to errors, time-consuming, which poses a barrier to the further success of deep learning. In [70], the authors presented an approach for segmenting nuclei from histopathological images that needs the least amount of annotation (i.e., centroid markings). The mixed-scale dense (MS-D) architecture [108], that underpins this method, allows for effective learning with a different cell sizes.

2.3   Conclusion

One of the main problems in histological image analysis is uncertainty. Some of the sources of this uncertainty include incompleteness and vagueness in stain class definition, as well as overlapping characteristics of histochemical stains. In this context, a novel circular clustering algorithm is introduced in the next chapter for color normalization of histological images, which judiciously integrates the merits of both fuzzy and rough sets to deal with the uncertainty associated with histological image analysis.

Chapter 3

Rough-Fuzzy Circular Clustering for Color Normalization of Histological Images

3.1 Introduction

Histology is defined as the microscopic study of tissue sections to analyze the manifestation of diseases under consideration. The most important property of histological data is the presence of enormous amount of information, more cellular details, compared to radiological , cytological and other imaging modalities. As a result, automated diagnosis based on histological data produces more accurate results than other imaging modalities. To ease pathologists' analysis, tissue samples collected from different biopsy sets are stained with several histochemical reagents, which in turn highlight different cellular structures. Staining not only highlights different cellular components, but also provides useful information in proper diagnosis of several pathological conditions [42]. Hence, in histological image analysis, color plays an important role in identification of different histological components. Computer-aided diagnosis can significantly eliminate imperfection associated with human interpretation in histological analysis of prostate, cervical and breast cancer diagnosis [56].

One of the most important problems in histological image analysis is the unacceptable variation in stained tissue color. The color representation of the stained histological images may vary due to various factors, such as manual sectioning of the tissue samples during specimen preparation, orientation of the lens aperture, inconsistency in staining procedure, variation in quality of stains obtained from different manufacturers, storage condition, inter-patient and inter-biopsy staining variations and so on. As a result, numerical features, extracted from these histological images with an observable variation in color representation, may be inaccurate and may lead to difficulty in image interpretation by computer-aided systems [15,76]. Hence, deviation in color representation affects proper detection, trustworthy diagnosis and gradation of the diseases under consideration. The color disagreement creates difficulty in image interpretation by software trained on a particular color appearance.

Color normalization is a process that reduces the inadmissible color variation corresponding to similar biological components within a single image or among the images within a data set maintaining the structural and histological information contained in the images. With the advancement of technology, some recent studies have ensured the importance of color features in quantitative analysis of histological images [69, 75]. As the color normalization methods try to reduce color deviation among the images, similar cellular structures in different images should exhibit similar color representation. Thus, the resulting images facilitate subsequent image segmentation and feature extraction, eventually improving the accuracy of the computer-aided diagnosis system. The most primary and simple approach in this domain is the histogram specification approach, which maps color representation of a source/query image to that of a pre-specified template image [20]. In another variation of histogram specification [116], color map quantile matching is used for color normalization. As histogram specification based methods do not investigate local neighborhood information as well as actual reasons associated with color variation problem, histological information is not expected to be kept intact after color normalization. Another group of color normalization approaches follows the color transfer technique, where each image is converted from RGB color space to $L\alpha\beta$ color space, as referred in [34], and corresponding to each image channel, mean and standard deviation are computed. Based on the source and template image channel statistics, source image is standardized [38].

Color normalization approach based on stain separation considers local information of the image content and tries to deal with the actual reasons associated with color variation among the images. To normalize the color of histological images, a non-linear Bezier curve-fitting based mapping approach is used in [15]. In this method, global stain color descriptor and local pixel-level information are fed to a relevance vector machine for classifying each pixel to a particular stain class. In [84], each RGB pixel is first transformed to corresponding optical density (OD) values and the OD values are projected onto a plane, spanned by the eigenvectors corresponding to two largest eigenvalues computed from singular value decomposition of the OD pixels. In [132], the histological images within the set are subjected to an illumination normalization module to get rid of color variation due to illumination inconsistency. After converting the RGB image to corresponding HSI counterpart, saturation-weighted hue histogram is computed, and finally k-means clustering is performed to accomplish the pixel clustering task. A major drawback of color deconvolution based normalization methods is their dependency on the availability of supervised control slides. In the plane fitting (PF) approach [84], a geometric solution has been proposed to achieve stain decomposition using singular value decomposition followed by thresholding. The major limitation of this approach is that the associated thresholds and control parameters are prefixed and cannot be updated adaptively according to image content, which may lead to faulty stain decomposition. A sparse non-negative matrix factorization (NMF) based blind stain estimation procedure is proposed in [22], where the imposed sparsity on the NMF objective function ensures the uniqueness of the solution. As these aforementioned methods do not consider the overlapping nature of histochemical stains, the color decomposition through stain vector estimation is supposed to be faulty.

One of the main problems in histological image analysis is uncertainty. This overlapping nature of multiple contrasting histochemical stains introduces the uncertainty in stain class definition. In [134], a circular mixture modeling based blind stain decomposition method is recently proposed, which considers the physical constraints such as non-negativity of the

decomposed matrices as well as overlapping property of the histochemical stains. It assumes that a single hue value, which is equal to the mean, ensures the definite belongingness of a pixel to the stain class, as it can only possess highest probability of belongingness to the stain class. However, a single hue value can never represent a stain class perfectly. In reality, a specific range or a set of hue values around mean represents the definite belongingness of pixels to a stain class. Each stain class should be defined using two sets: a core region, consisting of multiple hue values to represent definite belongingness of pixels, and an overlapping boundary region. In this background, rough sets [147] and fuzzy set [71], together may be effective to introduce the above concept while handling uncertainty present in histological image analysis.

In this regard, a new circular clustering algorithm, termed as rough-fuzzy circular clustering (RFCC$_{\text{cosine}}$), is proposed in this chapter for color normalization of histological images. The proposed algorithm assumes that each stain class is represented by a centroid, a crisp lower approximation, and a fuzzy boundary. While rough sets deal with the uncertainty, vagueness and incompleteness in stain class definition, fuzzy sets address the overlapping nature of histochemical reagents. The proposed method also integrates saturation-weighted hue histogram to reduce the impact of less-saturated achromatic pixels on stain color estimation. Some new quantitative measures are also proposed to evaluate the color constancy after normalization. The efficiency of the proposed method, along with a comparison with other state-of-the-art methods, is demonstrated on fifty-eight standard histological images. Some of the results presented in this chapter are reported in [104].

The structure of the rest of this chapter is as follows: Section 3.2 presents a new rough-fuzzy circular clustering algorithm, along with the estimation of associated parameters and other theoretical analysis, to deal with the stain color normalization problem of histological images. Section 3.3 introduces two new quantitative measures to evaluate between-image and within-set color constancy of region-of-interests after color normalization. The effectiveness of the proposed method is analyzed with respect to several state-of-the-art color normalization approaches on publicly available histological image data set in Section 3.4. Concluding remarks are drawn in Section 3.5.

3.2 Proposed Method

The stain color appearance map and stain density map, corresponding to each histology image, can be estimated using a stain decomposition method. In the proposed method, the stain decomposition is performed following the transmission light microscopy based image model [132]. Since the stain vectors intersect among each other at the origin in the OD domain, at first each image is mapped to OD space and then decomposed to extract the stain color appearance map and corresponding stain density map. The stain color appearance map or the spectrum matrix is estimated using a circular clustering, referred to as rough-fuzzy circular clustering, which follows next.

3.2.1 Rough-Fuzzy Circular Clustering

In the proposed rough-fuzzy circular clustering algorithm, each stain class is designated by a cluster prototype, a crisp lower approximation region, and a fuzzy boundary region. Let

us suppose that the lower approximation and boundary region, corresponding to the i-th stain class β_i, are represented by $\underline{A}(\beta_i)$ and $B(\beta_i)$, respectively. The upper approximation $\overline{A}(\beta_i)$ of stain class β_i is then defined by $\overline{A}(\beta_i) = \{\underline{A}(\beta_i) \cup B(\beta_i)\}$.

According to the notions of lower approximation and boundary region of rough sets, if the j-th hue x_j lies in the lower approximation $\underline{A}(\beta_i)$ of the i-th stain class β_i, then $x_j \notin \underline{A}(\beta_k), \forall k \neq i$ and also $x_j \notin B(\beta_i), \forall i$. That is, hue x_j certainly belongs to the stain class β_i. Hence, the weights of the hue values, belonging to the lower approximation region of a particular stain class, should be independent of other centroids and stain classes, and should not be coupled with their similarity with respect to other centroids. Also, the hue values lying in lower approximation of a particular stain class should have identical influence on respective centroid and stain class. On the other hand, if a hue value x_j belongs to the boundary region $B(\beta_i)$ corresponding to the i-th stain class β_i, then x_j possibly lies in β_i and potentially belongs to another stain class. Hence, the hue values lying in boundary regions should have different influence on the centroids and stain clusters.

To estimate the stain class representative vectors, the proposed rough-fuzzy circular clustering algorithm uses a saturation-weighted (SW) hue histogram H^{sw}, instead of standard hue histogram. It has been shown in [10] that, the standard hue histogram exhibits a number of sharp ridges, attributed by achromatic pixels. Thus, the hue values corresponding to the achromatic region are ill-defined and insignificant for the pixels possessing low saturation values. As a result, color distribution analysis, based on standard hue histogram, produces insignificant and erroneous results. The effects of the achromatic pixels on the standard hue histogram can be reduced by computing saturation-weighted hue histogram H^{sw} as follows:

$$H^{sw}(x) = \sum_{k \in I} s_k \delta(x, h_k); \quad \text{where} \quad \delta(x, h_k) = \begin{cases} 1 & \text{if } x = h_k \\ 0 & \text{otherwise.} \end{cases} \tag{3.1}$$

Here, I denotes a digital histological image. The hue and saturation values, corresponding to the k-th pixel in the HSI color space, are represented by h_k and s_k, respectively. The proposed rough-fuzzy circular clustering algorithm clusters the saturation-weighted hue histogram H^{sw}, corresponding to a histological image I, into c stain classes (an achromatic region and $(c-1)$ stain regions) by minimizing the following objective function, defined over a parameter set ψ:

$$J(\psi) = \sum_{i=1}^{c} \left[\omega \times J_i^L(\psi) + (1 - \omega) \times J_i^B(\psi) \right]; \tag{3.2}$$

where the parameters ω and $(1 - \omega)$ control relative importance between lower approximation and boundary region corresponding to each stain class, and

$$J_i^L(\psi) = \sum_{x_j \in \underline{A}(\beta_i)} [1 - \cos(x_j - v_i)] H^{sw}(x_j); \tag{3.3}$$

$$J_i^B(\psi) = \sum_{x_j \in B(\beta_i)} \mu_{ij}^m [1 - \cos(x_j - v_i)] H^{sw}(x_j). \tag{3.4}$$

Here, x_j represents the j-th hue and v_i denotes the i-th centroid corresponding to the

i-th stain class β_i, $m \in [1, \infty)$ represents the fuzzifier and $\mu_{ij} \in [0, 1]$, represents the fuzzy membership of the j-th hue into i-th stain class β_i. Hence, the parameter set ψ is given by $\psi = \{v_i, \mu_{ij}\}$.

Combining (3.2) and (3.4), the objective function $J^B(\psi)$, corresponding to the boundary regions, can be written as follows:

$$J^B(\psi) = \sum_{i=1}^{c} \sum_{x_j \in B(\beta_i)} \mu_{ij}^m [1 - \cos(x_j - v_i)] H^{sw}(x_j). \tag{3.5}$$

3.2.1.1 Fundamental Properties of Rough-Fuzzy Circular Clustering

From the formulation and discussion about the proposed rough-fuzzy circular clustering algorithm, the following properties can be stated:

1. $\underline{A}(\beta_i) \cap \underline{A}(\beta_k) = \varnothing, \forall i \neq k$.

2. $\underline{A}(\beta_i) \cap B(\beta_i) = \varnothing, \forall i$.

3. $\exists i, k, B(\beta_i) \cap B(\beta_k) \neq \varnothing$.

4. $\mu_{ij} = 1, \forall \theta_j \in \underline{A}(\beta_i)$.

5. $\mu_{ij} \in (0, 1), \forall \theta_j \in B(\beta_i)$.

The property 1 states that if a hue value θ_j belongs to $\underline{A}(\beta_i)$, it implies that $\theta_j \notin \underline{A}(\beta_k), \forall k \neq i$, that is, the hue θ_j is an element of the stain class β_i definitely. The property 2 ensures that if $\theta_j \in \underline{A}(\beta_i)$, then $\theta_j \notin B(\beta_i)$, that is, a hue cannot belong to both lower and boundary regions of a stain class β_i. The property 3 says that $\theta_j \in B(\beta_i)$ implies that $\exists k$ such that $\theta_j \in B(\beta_k)$. It means a hue $\theta_j \in B(\beta_i)$ possibly belongs to β_i and potentially belongs to other stain classes. The properties 4 and 5 are essential in computing the objective function $J^B(\psi)$ and the cluster prototype μ. They say that the membership value of the hue in lower approximation is 1, while for boundary region it remains unchanged. That is, each stain class β_i contains a crisp lower approximation $\underline{A}(\beta_i)$ and a fuzzy boundary $B(\beta_i)$.

3.2.1.2 Estimation of Membership Function

To optimization the objective function $J^B(\psi)$ with respect to μ_{ij}, subject to the constraint $\sum_{i=1}^{c} \mu_{ij} = 1, \forall j$, $J^B(\psi)$ is transformed to the following Lagrangian function:

$$\mathcal{J}^B(\psi) = \sum_{i=1}^{c} \sum_{x_j \in B(\beta_i)} \mu_{ij}^m [1 - \cos(x_j - v_i)] H^{sw}(x_j) + \lambda \sum_{x_j \in B(\beta_i)} \left[1 - \sum_{i=1}^{c} \mu_{ij} \right] H^{sw}(x_j); \tag{3.6}$$

where λ denotes Lagrange's multiplier. To estimate the fuzzy membership μ_{ij} of the j-th hue into i-th stain class β_i, $\mathcal{J}^B(\psi)$ is partially differentiated with respect to μ_{ij} as follows:

$$\frac{\partial \mathcal{J}^B(\psi)}{\partial \mu_{ij}} = 0;$$

$$\Rightarrow m\mu_{ij}^{m-1} d_{ij} H^{sw}(x_j) - \lambda H^{sw}(x_j) = 0 \Rightarrow \mu_{ij} = \left(\frac{\lambda}{md_{ij}}\right)^{\frac{1}{m-1}}. \tag{3.7}$$

Here, $d_{ij} = [1 - \cos(x_j - v_i)]$ represents the cosine distance. So,

$$\sum_{i=1}^{c} \mu_{ij} = 1 \Rightarrow \left(\frac{\lambda}{m}\right)^{\frac{1}{m-1}} = \left(\sum_{k=1}^{c} \left(\frac{1}{d_{kj}}\right)^{\frac{1}{m-1}}\right)^{-1}. \tag{3.8}$$

Substituting (3.8) in (3.7), the membership function μ_{ij} can be estimated as follows:

$$\mu_{ij} = \left(\frac{\lambda}{m}\right)^{\frac{1}{m-1}} \left(\frac{1}{d_{ij}}\right)^{\frac{1}{m-1}}; \quad \Rightarrow \mu_{ij} = \left(\sum_{k=1}^{c} \left(\frac{d_{ij}}{d_{kj}}\right)^{\frac{1}{m-1}}\right)^{-1}. \tag{3.9}$$

It is clearly observed from (3.9) that the computation of membership function μ_{ij} is dependent on the fuzzifier value m.

3.2.1.3 Computation of Cluster Prototype

The computation of cluster prototype depends on the weighted average of crisp lower approximation and fuzzy boundary region corresponding to each stain class. The effects of both fuzzy membership, and lower approximation and boundary regions are incorporated in the computation of the centroid. The centroid v_i, corresponding to the i-th stain class β_i, is computed by differentiating (3.2) partially with respect to v_i, as follows:

$$\frac{\partial J(\psi)}{\partial v_i} = 0;$$

$$\Rightarrow \omega \times \sum_{x_j \in \underline{A}(\beta_i)} \sin(x_j - v_i) H^{sw}(x_j) + (1 - \omega) \times \sum_{x_j \in B(\beta_i)} \mu_{ij}^{m} \sin(x_j - v_i) H^{sw}(x_j) = 0;$$

$$\Rightarrow v_i = \arctan\left[\frac{\omega \times \mathcal{P}_1 + (1 - \omega) \times \mathcal{P}_2}{\omega \times \mathcal{Q}_1 + (1 - \omega) \times \mathcal{Q}_2}\right]; \tag{3.10}$$

where $\mathcal{P}_1 = \sum_{x_j \in \underline{A}(\beta_i)} \sin(x_j) H^{sw}(x_j), \mathcal{P}_2 = \sum_{x_j \in B(\beta_i)} \mu_{ij}^{m} \sin(x_j) H^{sw}(x_j);$

$$\mathcal{Q}_1 = \sum_{x_j \in \underline{A}(\beta_i)} \cos(x_j) H^{sw}(x_j), \text{ and } \mathcal{Q}_2 = \sum_{x_j \in B(\beta_i)} \mu_{ij}^{m} \cos(x_j) H^{sw}(x_j).$$

It is evident from (3.10) that the computation of centroid depends on selection of the parameter ω. As the hue values corresponding to lower approximation region certainly belong to a specific stain class, they are supposed to be assigned a higher weightage ω compared to a lower weightage $(1 - \omega)$, assigned to the hues belonging to the boundary region.

3.2.2 Proposed Algorithm and Color Normalization

The optimization of $J(\psi)$ using the proposed rough-fuzzy circular clustering algorithm is based on an alternating optimization method, referred as Picard iteration, through (3.9) and (3.10). The procedure starts by initializing the parameters $v_i^{(0)}$ and $\mu_{ij}^{(0)}$ corresponding to $(c-1)$ stain regions and an achromatic region. Depending on the circular nature of hue, circular thresholding method of [139] is utilized to produce approximate partitions corresponding to the achromatic region and different stain regions. The initial cluster prototype $v_i^{(0)}$, corresponding to both stain and achromatic regions, can be computed from the saturation-weighted hue histogram H^{sw} using the initial parameter estimation method, proposed in [134]. The fuzzy memberships, associated to all hues, are then computed using (3.9). The basic steps of the proposed rough-fuzzy circular clustering algorithm are outlined as follows:

Algorithm 3.1 Proposed rough-fuzzy circular clustering algorithm.

Input: Saturation-weighted hue histogram H^{sw}, number of stain classes c.
Output: Estimated parameter set $\psi = \{v_i, \mu_{ij}\}$.

1: Initialize $v_i^{(0)}$ $(1 \leqslant i \leqslant c-1)$ for $(c-1)$ stain regions and $v_{ac}^{(0)}$ for the achromatic region.
2: Select values for fuzzifier m, and pre-specified thresholds ϵ and δ. Set iteration counter $\kappa = 1$.
3: Compute μ_{ij} for c classes and n hues using (3.9).
4: If μ_{ij} and μ_{kj} represent the highest and second highest memberships of x_j, respectively, and $(\mu_{ij} - \mu_{kj}) > \delta$, then $x_j \in \underline{A}(\beta_i)$; otherwise $x_j \in B(\beta_i)$ and $x_j \in B(\beta_k)$.
5: Update μ_{ij} considering the lower approximation and boundary regions for c classes and n hues.
6: Compute new centroid v_i as per (3.10).
7: Repeat steps 3 to 6, by incrementing κ, until $|\mu_{ij}^{(\kappa)} - \mu_{ij}^{(\kappa-1)}| < \epsilon$.
8: Stop.

The spectrum matrix can be computed from the cluster centroids representing each stain class. The cluster prototype, v_i, $i = 1, 2, \cdots, c$, estimated from the proposed circular clustering algorithm, is considered to be the representative hue h_i for the i-th stain class β_i. The saturation s_i and intensity value y_i, corresponding to the representative hue h_i, are computed using the saturation-weighted statistics, described in [132]. The HSI domain representative stain vector $I_i^{hsi} = [h_i, s_i, y_i]^t$, corresponding to the i-th stain class β_i, is first transformed to RGB color space: $I_i^{rgb} = G_{hsi}^{rgb}(I_i^{hsi})$, where the function $G_{hsi}^{rgb}(.)$ converts a HSI domain stain vector to corresponding RGB domain vector. Thus, the stain vectors corresponding to the c stain classes are utilized to form the representative spectrum matrix of dimension $3 \times c$. The RGB domain color appearance vector I_i^{rgb} is then converted to corresponding OD domain vector represented as: $W_i = \log(I^b/I_i^{rgb})$, where I^b denotes the intensity value corresponding to a background pixel. The stain density value corresponding to a pixel p can be evaluated using the following equation:

$$X(p) = W^{-1}[\log(I^b/I(p))]. \tag{3.11}$$

As the spectrum matrix and corresponding stain density map should satisfy the element-
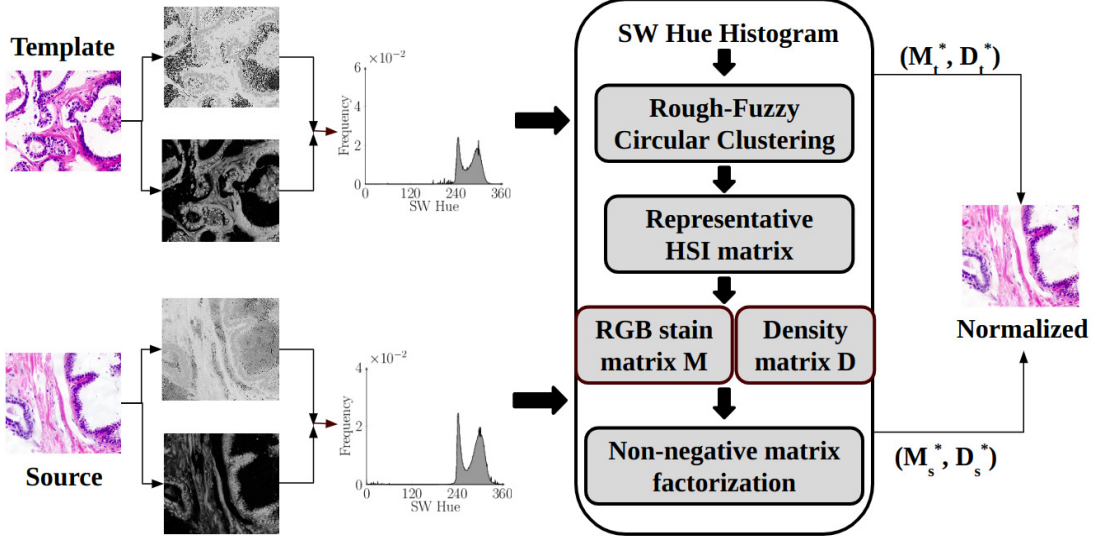
Figure 3.1: Block diagram of the proposed color normalization method

wise non-negativity constraint, the NMF is applied to refine the decomposition solution:

$$\{W^*, X^*\} = \min_{W \geq 0, X \geq 0} \| \log(I^b/I) - W \times X \|^2 . \tag{3.12}$$

Here, ≥ 0 represents element-wise non-negativity corresponding to the factor matrices. Using the template image stain color appearance map W^*, the source image color is normalized maintaining the stain density information X^* associated with each pixel of the source image. The block diagram of the proposed rough-fuzzy circular clustering based color normalization method is depicted in Figure 3.1.

3.2.3 Convergence Analysis

In this section, a mathematical analysis on the convergence criteria of the proposed rough-fuzzy circular clustering algorithm is provided. To examine the convergence condition of the proposed algorithm, the Hessian (second order partial derivative) matrices with respect to the parameters involved in the objective function $J(\psi)$ are computed and then checked whether the Hessian matrices are positive definite or not. The second order partial derivative of the objective function defined in (3.2) with respect to the cluster centroid v_i leads to the following

$$\frac{\partial^2 J(\psi)}{\partial \mu_i^2} = \omega \times M_i + (1 - \omega) \times N_i; \tag{3.13}$$

$$\text{where} \quad M_i = \sum_{x_j \in \underline{A}(\beta_i)} \cos(x_j - v_i) H^{sw}(x_j); \tag{3.14}$$

$$\text{and} \quad N_i = \sum_{x_j \in B(\beta_i)} \mu_{ij}^m \cos(x_j - v_i) H^{sw}(x_j). \tag{3.15}$$

28

It is evident that $\mu_{ij} \geqslant 0$ implies $\mu_{ij}^m \geqslant 0$, $\cos(x_j - v_i) \geqslant 0$ and $H^{sw}(x_j) \geqslant 0$ $\forall x_j$. Hence, for each cluster β_i, $M_i \geqslant 0$ and $N_i \geqslant 0$. Also, it can be easily verified that, $\forall k \neq i$,

$$\frac{\partial^2 J(\psi)}{\partial v_k \partial v_i} = 0. \tag{3.16}$$

Hence, it is clearly noticed from (3.16) that all the off-diagonal entries in the respective Hessian matrices are 0. The proposed algorithm is guaranteed to converge if the Hessian matrix with respect to v_i is diagonally dominant [2]. This is a sufficient condition, but not a necessary one. If the Hessian matrix is not diagonally dominant, the iteration may or may not converge. The Hessian matrix corresponding to (3.14) is given by $\mathcal{X} = \mathrm{diag}\{M_1, \cdots, M_i, \cdots, M_c\}$ and the matrix corresponding to (3.15) is given by $\mathcal{Y} = \mathrm{diag}\{N_1, \cdots, N_i, \cdots, N_c\}$.

It can be easily observed that the Hessian matrices $\mathcal{X}$ and $\mathcal{Y}$ are diagonally dominant. So, $\mathcal{X}$ and $\mathcal{Y}$ are certainly positive definite matrices. Hence, all the eigenvalues corresponding to $\mathcal{X}$ and $\mathcal{Y}$ are positive. Under this condition, the iterative algorithm would converge to at least a local optimum solution if (3.10) is repetitively applied keeping μ_{ij} constant. Similarly, it can be easily proved that the Hessian matrices with respect to membership function μ_{ij} is diagonal positive definite matrices. Intuitively, the objective function $J(\psi)$ reduces in all steps corresponding to (3.9) and (3.10), which makes the system strictly descent. The above analysis ensures the convergence of the proposed rough-fuzzy circular clustering algorithm.

3.3 Quantitative Measures

One of the important criteria of color normalization is to reduce the within-image as well as between-image color variation of a single biopsy set for a particular region of interest (ROI). It is assumed that the tissue specimens within the same biopsy set have undergone same staining treatment and thus they must exhibit similar color distribution. Though, normalized median intensity (NMI) [74] evaluates the color consistency of a specific ROI within an image, it does not capture the color consistency of the ROI among images within the same biopsy set. In order to address the above problem, two new quantitative indices are introduced next to evaluate the performance of different color normalization methods.

3.3.1 Between-Image Color Constancy Index

The between-image color constancy (BiCC) index of an image I, belonging to a biopsy set S, is defined as follows:

$$\mathrm{BiCC}(I) = \frac{1}{2(|S|-1)} \times \sum_{J \neq I} \left\{ \frac{\underset{i \in \mathrm{ROI}(I)}{\mathrm{median}}\{W(i)\} + \underset{j \in \mathrm{ROI}(J)}{\mathrm{median}}\{W(j)\}}{\max\left\{\underset{i \in \mathrm{ROI}(I)}{\max}\{W(i)\}, \underset{j \in \mathrm{ROI}(J)}{\max}\{W(j)\}\right\}} \right\}; \tag{3.17}$$

where $W(i)$ denotes the average of (R, G, B) intensities for the i-th pixel corresponding to the ROI of image I. If the between-image color constancy is maintained after color

29

normalization, it is expected that the median and maximum average intensity values, corresponding to ROIs of images I and J, would exhibit a close proximity. So, a good color normalization algorithm should make the value of BiCC as high as possible. The value of BiCC lies in between 0 and 1.

3.3.2   Within-Set Color Constancy Index

The within-set color constancy (WsCC) index for a biopsy set S is defined as follows:

$$\text{WsCC}(S) = \frac{1}{|S|} \sum_{I \in S} \text{NMI}(I) \times \text{BiCC}(I); \qquad (3.18)$$

where $\text{NMI}(I)$ is the normalized median intensity of image I and $\text{BiCC}(I)$ is the between-image color constancy of image I with respect to the biopsy set S. NMI is defined as follows:

$$\text{NMI}(I) = \frac{\underset{i \in \text{ROI}(I)}{\text{median}}\{W(i)\}}{\underset{i \in \text{ROI}(I)}{\max}\{W(i)\}}. \qquad (3.19)$$

So, if the color constancy is maintained after normalization, then it is expected that the median and maximum of the average intensity values, corresponding to a particular ROI of image I, would be close enough; which indicates a high value of NMI. Also, the value of NMI ranges from 0 to 1. So, a good color normalization algorithm should make the WsCC for a biopsy set S high as both NMI and BiCC of a particular image, belonging to the biopsy set S, eventually possess high values. The value of WsCC varies in between 0 and 1.

3.4   Experimental Results and Discussions

The performance of the proposed rough-fuzzy circular clustering based stain separation method is studied extensively and compared with that of

- several state-of-the-art stain separation methods, such as, plane fitting (PF) [84], HTN method as referred in [132], structure-preserving color normalization (SPCN) [22] and expectation-maximization (EM) based circular mixture modeling [134]; and

- different existing color normalization methods, namely, color transfer (CT) technique [38], stain color description (SCD) [15] as well as PF [84], HTN [132], and SPCN [22].

As the proposed rough-fuzzy circular clustering algorithm utilizes cosine distance as the distance measure, the proposed algorithm is referred to as RFCC$_{\text{cosine}}$ in this chapter and further chapters.

3.4.1   Description of Data Sets

To assess the performance of several stain decomposition and color normalization approaches, 58 images from publicly available histology image data set UCSB is considered

in this chapter. This UCSB data set, published by University of California, Santa Barbara [36], consists of 32 non-cancerous benign cell and 26 cancerous malignant cell images, collected from 10 hematoxylin (H) and eosin (E) stained breast cancer biopsy sets. As UCSB data set comprises of images stored in 24 bit nonlinear RGB format, each image in this set is transformed to linear RGB space before analysis [68]. Each image has dimension 896×768.

To find out the optimum values of weight parameter ω for the proposed method, the value of ω is varied from 0.51 to 0.99. The optimum values of ω for a particular image, is obtained using Silhouette index [103].
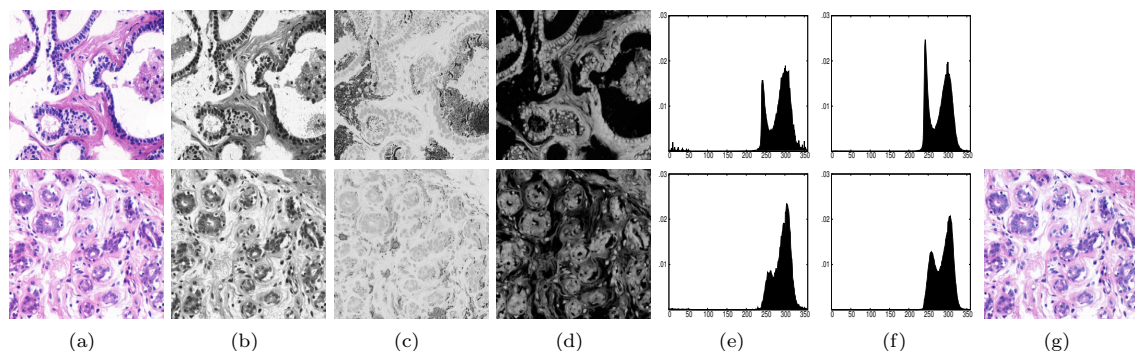
Figure 3.2: Example of color normalization (top: template; bottom: source): (a) original; (b) intensity; (c) hue; (d) saturation; (e) hue histogram; (f) saturation-weighted hue histogram; (g) color normalized source

3.4.2 Illustrative Example

Figure 3.2 presents an illustrative example of the proposed color normalization algorithm, considering ytma23_022103_benign2_ccd of biopsy set ytma23_022103 as template and ytma10_010704_benign1_ccd of biopsy set ytma10_010704 as query/source image, respectively from the UCSB data set. While Figure 3.2(a) represents the original image of both template and source, Figures 3.2(b)-(d) depict the corresponding intensity, hue, and saturation, respectively, of both images. Figures 3.2(e) and (f) show the hue histograms and corresponding saturation-weighted hue histograms. Comparing Figures 3.2(e) and (f), it can be seen that the saturation-weighted hue histogram can eliminate the sharp ridges, which are mainly attributed by achromatic pixels and exhibited in standard hue histogram. Three cluster centroids obtained using the proposed rough-fuzzy circular clustering algorithm corresponding to the source image, considering saturation-weighted hue histogram, are $v_{ac} = 20.907$, $v_H = 262.089$, and $v_E = 303.857$. Figure 3.2(g) depicts the color normalized source image obtained using the proposed color normalization algorithm.

3.4.3 Relevance of Saturation-Weighted Hue Histogram

In order to ensure the usefulness of saturation-weighted hue histogram over the standard hue histogram, extensive analysis is performed on UCSB data set, and corresponding results are reported in Table 3.1. The UCSB data set consists of 10 biopsy sets: 9 of them

Table 3.1: Performance of Standard and Saturation-Weighted Histograms on UCSB Data

Different Biopsy Sets	Different Channels	Standard Histogram		SW Histogram	
		σ_{H}	σ_{E}	σ_{H}	σ_{E}
ytma10_010704	Red	6.31	**2.68**	**2.39**	3.65
	Green	3.80	1.08	**1.39**	**1.03**
	Blue	2.22	**2.20**	**1.59**	2.55
ytma12_010804	Red	3.99	**5.26**	**1.89**	5.73
	Green	3.16	**1.63**	**1.17**	2.03
	Blue	1.25	1.46	**0.94**	**1.19**
ytma23_022103	Red	**1.61**	**5.12**	1.67	5.31
	Green	1.33	**1.53**	**1.24**	1.80
	Blue	0.69	1.49	**0.64**	**1.38**
ytma49_042003	Red	3.98	3.07	**3.28**	**2.36**
	Green	2.41	**0.45**	**1.99**	0.54
	Blue	1.02	1.51	**0.69**	**0.94**
ytma49_042203	Red	7.53	5.77	**2.24**	**3.07**
	Green	4.09	5.90	**1.49**	**0.87**
	Blue	1.65	12.86	**0.64**	**1.89**
ytma49_042403	Red	2.28	**3.29**	**1.72**	3.80
	Green	1.38	**0.91**	**0.94**	1.10
	Blue	**0.71**	1.28	0.76	**0.72**
ytma49_072303	Red	1.45	4.24	**1.41**	**2.86**
	Green	**1.14**	1.19	**1.14**	**1.09**
	Blue	**0.29**	1.08	0.30	**0.67**
ytma49_111003	Red	**1.89**	**3.24**	2.79	5.34
	Green	**1.51**	**1.31**	2.39	2.69
	Blue	**1.01**	**1.44**	1.08	1.93
ytma49_111303	Red	**1.70**	6.29	2.29	**6.09**
	Green	**1.43**	**2.44**	2.06	2.54
	Blue	**0.79**	1.82	0.80	**1.80**
ytma55_030603	Red	4.18	6.46	**1.91**	**3.84**
	Green	3.21	3.57	**1.52**	**1.62**
	Blue	0.80	10.41	**0.77**	**1.67**
Note: σ_{X} represents the standard deviation of the estimated stain vectors of X-stain, corresponding to all images of the same biopsy set; X $\in$ {H, E}					

contain 6 images each, whereas another contains only 4 images. It can be assumed that the images within the same biopsy set have undergone similar type of staining treatment. As a result, the images within the same biopsy set are expected to produce similar or consistent spectrum vectors. Based on this intuition, the standard deviation (σ) values of the estimated stain vectors of all the images, corresponding to the same biopsy set, are summarized in Table 3.1 (values are represented in 10^{-2} scale). As the spectrum vectors, corresponding to the same biopsy set, must exhibit low standard deviation, the smallest σ value, associated to respective matrix element, is marked in bold. It can be observed from Table 3.1 that the saturation-weighted hue histogram attains lowest σ values in 21 and 17 cases, out of 30 cases each for H-stain and E-stain, respectively, while the standard hue histogram achieves lowest σ in only 10 and 13 cases.

Moreover, the consistency of the color distribution, after color normalization, is evaluated with respect to normalized median intensity (NMI) [74] measure. The comparative performance analysis of the proposed method and standard histogram, and hard and fuzzy

Table 3.2: Comparison of Different Histograms and Clustering Algorithms Using NMI

Different Histograms /Clustering Methods	Normalized Median Intensity (NMI)			
	Mean	Median	Wilcoxon Test	Paired-t Test
Standard Histogram	0.700852	0.697148	9.29E-09	2.42E-06
Hard Clustering	0.701822	0.696991	3.49E-09	1.22E-07
Fuzzy Clustering	0.702271	0.696991	9.48E-09	2.18E-07
RFCC$_{cosine}$	**0.710847**	**0.707624**		

Table 3.3: Comparison of Different Histograms and Clustering Algorithms Using BiCC

Different Histograms /Clustering Methods	Between-image Color Constancy (BiCC) index			
	Mean	Median	Wilcoxon Test	Paired-t Test
Standard Histogram	0.670558	0.657639	1.75E-11	2.19E-17
Hard Clustering	0.670490	0.658078	1.75E-11	2.28E-19
Fuzzy Clustering	0.671444	0.658307	1.75E-11	2.13E-18
RFCC$_{cosine}$	**0.692504**	**0.681332**		

clustering based counterparts is performed by means of tables of mean, median, and p-values computed using both paired-t (one-tailed) and Wilcoxon signed rank (one-tailed) tests. From the results reported in Table 3.2, it can be observed that the saturation-weighted hue histogram produces highest mean and median values for NMI and performs significantly better than the standard hue histogram, considering 0.05 as the level of significance. The between-image and within-set color constancy of the proposed RFCC$_{cosine}$ method after color normalization are compared with that of standard histogram and clustering based variants of the model with respect to introduced BiCC and WsCC index and the comparative results are provided in Table 3.3 and 3.4, respectively. It can be easily observed from the results reported in Table 3.3 and 3.4 that the proposed method performs better than its variants as per as color constancy after normalization is concerned.

3.4.4 Significance of Rough Sets

To establish the significance of the proposed rough-fuzzy circular clustering algorithm over its hard and fuzzy counterparts, the results on UCSB data set are reported in Table 3.5. It can be seen from the results of Table 3.5 that the proposed rough-fuzzy circular clustering algorithm achieves lowest σ values in 15 and 18 cases for H-stain and E-stain, respectively, while hard and fuzzy clustering attain lowest σ in 10 and 14, and 13 and 3 cases, respectively. Moreover, the proposed circular clustering algorithm provides significantly better NMI values, compared to both hard and fuzzy clustering as observed in Table 3.2. The proposed rough-fuzzy circular clustering algorithm performs significantly better than its hard and fuzzy counterparts as the incorporation of rough sets can efficiently handle uncertainty in stain class definition due to incompleteness and vagueness.

3.4.5 Performance of Different Algorithms

In order to establish the efficacy of the proposed rough-fuzzy circular clustering based stain estimation method, the performance of the proposed algorithm is compared with that of

Table 3.4: Comparison of Different Histograms and Clustering Algorithms Using WsCC

Different Histograms /Clustering Methods	Within-set Color Constancy (WsCC) index			
	Mean	Median	Wilcoxon Test	Paired-t Test
Standard Histogram	0.474186	0.465024	1.75E-11	2.40E-16
Hard Clustering	0.473908	0.464562	1.75E-11	3.54E-18
Fuzzy Clustering	0.475167	0.465122	1.75E-11	3.44E-17
RFCC$_{\text{cosine}}$	**0.494584**	**0.475552**		

Table 3.5: Comparative Performance of Different Clustering Algorithms on UCSB Data

Different Biopsy Sets	Different Channels	Hard		Fuzzy		RFCC$_{\text{cosine}}$	
		σ_{H}	σ_{E}	σ_{H}	σ_{E}	σ_{H}	σ_{E}
ytma10_010704	Red	**2.37**	3.69	2.51	3.79	2.39	**3.65**
	Green	**1.25**	1.04	1.35	1.09	1.39	**1.03**
	Blue	1.26	2.52	**1.21**	**2.49**	1.59	2.55
ytma12_010804	Red	2.00	5.78	**1.89**	5.85	**1.89**	**5.73**
	Green	1.24	2.06	1.18	2.08	**1.17**	**2.03**
	Blue	0.95	1.20	**0.91**	1.21	0.94	**1.19**
ytma23_022103	Red	1.70	**5.30**	**1.67**	5.42	**1.67**	5.31
	Green	1.26	1.81	1.27	1.83	**1.24**	**1.80**
	Blue	**0.63**	**1.36**	0.66	1.39	0.64	1.38
ytma49_042003	Red	3.29	**2.36**	3.45	2.37	**3.28**	**2.36**
	Green	**1.99**	0.55	2.11	**0.54**	**1.99**	**0.54**
	Blue	**0.66**	**0.94**	0.94	0.95	0.69	**0.94**
ytma49_042203	Red	2.25	3.09	2.39	3.07	**2.24**	**3.07**
	Green	**1.44**	0.90	1.65	**0.87**	1.49	**0.87**
	Blue	0.72	1.80	**0.64**	1.90	**0.64**	1.89
ytma49_042403	Red	1.73	3.93	1.73	4.08	**1.72**	**3.80**
	Green	0.96	1.13	**0.94**	1.17	**0.94**	**1.10**
	Blue	0.76	0.77	**0.70**	0.87	0.76	**0.72**
ytma49_072303	Red	**1.41**	**2.83**	**1.41**	2.89	**1.41**	2.86
	Green	**1.13**	**1.08**	1.15	1.11	1.14	1.09
	Blue	0.31	0.69	0.32	0.73	**0.30**	**0.67**
ytma49_111003	Red	2.81	**5.26**	**2.67**	5.53	2.79	5.34
	Green	2.37	**2.64**	**2.32**	2.82	2.39	2.69
	Blue	1.13	**1.92**	1.10	1.97	**1.08**	1.93
ytma49_111303	Red	2.28	**5.96**	**2.25**	6.17	2.29	6.09
	Green	**2.04**	**2.50**	2.05	2.59	2.06	2.54
	Blue	**0.78**	**1.79**	0.84	1.82	0.80	1.80
ytma55_030603	Red	1.94	3.86	**1.85**	4.08	1.91	**3.84**
	Green	1.54	**1.62**	**1.51**	1.72	1.52	**1.62**
	Blue	0.78	1.71	0.78	1.73	**0.77**	**1.67**
Note: σ_{X} represents the standard deviation of the estimated stain vectors of X-stain, corresponding to all images of the same biopsy set; X $\in$ {H, E}							

different state-of-the-art stain separation methods, namely, PF [84], HTN [132], SPCN [22], and EM [134] and the results are summarized in Table 3.6. It is observed from Table 3.6 that the proposed method achieves lowest standard deviation (σ) values in 11 and 12 cases, out of 30 cases each, corresponding to H-stain and E-stain, respectively. On the other hand, the existing PF, HTN, SPCN, EM exhibit lowest σ values in 1 and 7, 8 and 6, 3 and 4, and 7 and 1 cases, respectively, corresponding to H-stain and E-stain. The results reported in

Table 3.6: Comparative Performance Analysis of Different Algorithms on UCSB Data

Different Biopsy Sets	Different Channels	PF		HTN		SPCN		EM		RFCC$_{cosine}$	
		σ_H	σ_E	σ_H	σ_E	σ_H	σ_E	σ_H	σ_E	σ_H	σ_E
ytma10_010704	Red	3.04	**2.72**	2.67	3.25	2.42	3.93	2.45	4.10	**2.39**	3.65
	Green	2.23	2.01	1.68	0.97	1.66	0.73	1.61	**0.69**	**1.39**	1.03
	Blue	2.67	5.79	1.14	**2.43**	1.01	2.55	**0.92**	2.51	1.59	2.55
ytma12_010804	Red	1.99	**3.79**	**1.37**	5.79	2.20	6.26	2.08	7.73	1.89	5.73
	Green	2.16	2.82	**1.07**	**1.82**	1.80	2.14	1.66	2.54	1.17	2.03
	Blue	1.72	6.81	**0.55**	1.30	0.79	1.36	0.83	1.56	0.94	**1.19**
ytma23_022103	Red	1.77	**3.43**	2.39	5.36	**1.07**	5.75	1.18	7.27	1.67	5.31
	Green	1.95	**0.46**	2.25	1.59	**1.10**	1.81	1.16	2.21	1.24	1.80
	Blue	1.31	3.34	0.70	1.57	0.86	1.71	0.89	1.86	**0.64**	**1.38**
ytma49_042003	Red	**3.11**	2.11	3.17	**1.86**	3.14	2.98	4.56	4.53	3.28	2.36
	Green	2.98	2.02	2.35	**0.45**	2.44	0.74	3.40	0.96	**1.99**	0.54
	Blue	1.09	3.67	**0.55**	0.98	0.79	1.18	1.21	1.42	0.69	**0.94**
ytma49_042203	Red	2.29	**2.18**	4.04	3.13	1.99	3.36	**1.93**	3.88	2.24	3.07
	Green	2.95	1.40	3.81	0.92	2.02	**0.84**	2.06	1.18	**1.49**	0.87
	Blue	1.88	3.72	1.15	2.12	**0.60**	2.03	0.69	2.02	0.64	**1.89**
ytma49_042403	Red	3.08	**3.13**	3.42	3.67	2.99	4.89	3.39	6.66	**1.72**	3.80
	Green	3.16	3.18	2.55	**0.96**	2.38	1.38	2.72	1.67	**0.94**	1.10
	Blue	1.04	5.38	**0.63**	**0.47**	0.93	0.64	0.92	0.80	0.76	0.72
ytma49_072303	Red	2.09	2.63	2.49	2.82	1.99	**1.57**	2.52	1.99	**1.41**	2.86
	Green	2.48	1.22	2.55	0.95	2.15	**0.66**	2.67	0.95	**1.14**	1.09
	Blue	0.57	1.43	0.38	0.80	0.46	0.79	0.36	0.98	**0.30**	**0.67**
ytma49_111003	Red	1.90	13.17	1.64	6.94	4.00	9.65	**1.62**	8.12	2.79	**5.34**
	Green	2.60	11.29	1.61	3.49	1.69	5.79	**1.45**	4.63	2.39	**2.69**
	Blue	3.62	10.73	1.23	2.79	9.77	4.17	1.29	2.62	**1.08**	**1.93**
ytma49_111303	Red	2.15	6.12	**1.20**	6.71	1.44	6.85	1.63	8.87	2.29	**6.09**
	Green	2.52	**1.75**	**1.30**	2.46	1.56	2.85	1.75	3.27	2.06	2.54
	Blue	1.50	4.30	**0.24**	2.34	0.34	2.31	0.55	2.70	0.80	**1.80**
ytma55_030603	Red	2.18	7.85	1.88	16.56	1.35	5.22	**1.25**	7.60	1.91	**3.84**
	Green	3.39	3.90	1.81	2.09	1.48	2.36	**1.37**	3.19	1.52	**1.62**
	Blue	2.73	8.20	0.67	8.32	0.70	**1.62**	**0.67**	2.28	0.77	1.67
Note: σ_X represents the standard deviation of the estimated stain vectors of X-stain, corresponding to all images of the same biopsy set; X $\in$ {H, E}											

Table 3.6 ensure that the proposed algorithm outperforms the state-of-the-art blind stain separation and stain estimation methods.

Table 3.7: Comparison of Different Color Normalization Algorithms Using NMI

Different Methods	Normalized Median Intensity (NMI)			
	Mean	Median	Wilcoxon Test	Paired-t Test
CT	0.661974	0.680989	4.83E-06	3.40E-06
PF	0.689482	0.689037	8.24E-04	2.05E-04
SCD	0.602906	0.586130	2.47E-10	3.81E-14
HTN	0.706873	0.701627	2.58E-04	1.65E-03
SPCN	0.680182	0.679416	4.49E-06	5.92E-07
RFCC$_{cosine}$	**0.710847**	**0.707624**		

The color consistency within the normalized images is next evaluated for the proposed method as well as other color normalization methods using the NMI index, and correspond-

Table 3.8: Comparison of Different Color Normalization Algorithms Using BiCC

Different Methods	Between-image Color Constancy (BiCC) Index			
	Mean	Median	Wilcoxon Test	Paired-t Test
CT	0.613403	0.625624	4.94E-11	1.09E-13
PF	0.657836	0.660921	8.73E-07	3.49E-08
SCD	0.566423	0.563674	1.75E-11	3.22E-22
HTN	0.674506	0.661148	2.16E-11	2.24E-14
SPCN	0.647584	0.646141	1.81E-08	1.76E-10
RFCC$_{cosine}$	**0.692504**	**0.681332**		

Table 3.9: Comparison of Different Color Normalization Algorithms Using WsCC

Different Methods	Within-set Color Constancy (WsCC) Index			
	Mean	Median	Wilcoxon Test	Paired-t Test
CT	0.409771	0.424711	5.91E-09	3.18E-10
PF	0.455051	0.455632	2.29E-05	2.50E-06
SCD	0.350092	0.329816	1.85E-11	6.50E-20
HTN	0.479276	0.470083	1.36E-10	4.90E-12
SPCN	0.441564	0.439264	1.52E-07	9.00E-09
RFCC$_{cosine}$	**0.494584**	**0.475552**		

ing results are reported in Table 3.7. The between-image and within-set color constancy among the histological images is analyzed using the two introduced indices, BiCC and WsCC, and the corresponding results are reported in Table 3.8 and 3.9, respectively. From the results reported in Table 3.7, 3.8 and 3.9 it is evident that the proposed method produces highest mean and median values as compared to other color normalization methods. Moreover, the paired-t and Wilcoxon signed rank tests ensure that the proposed method performs significantly better than the existing color normalization approaches. The qualitative comparisons of different color normalization methods are presented in Figure 3.3 and Figure 3.4. The results reported in Table 3.7, Figure 3.3 and Figure 3.4 establish the fact that the proposed method outperforms state-of-the-art color normalization methods as per the color consistency after color normalization is concerned.

3.5 Conclusion

In histological image analysis, color normalization is one of the most important and foremost tasks, as the presence of color disagreement among the images may affect the performance of automated histological image analysis. In this regard, the contributions of this chapter are mainly two-fold: (i) the development of a new circular clustering algorithm, designed for color normalization of histological images, which judiciously integrates the merits of rough sets and fuzzy set theory; and (ii) demonstrating the efficiency of the proposed method, along with a comparison with other state-of-the-art methods, on standard H&E stained histological image set. The proposed method deals with the overlapping nature of stain classes as well as addresses the non-negativity constraint corresponding to stain color appearance map and corresponding stain density map. From the reported quantitative and qualitative results, it is clear that the proposed method outperforms existing stain
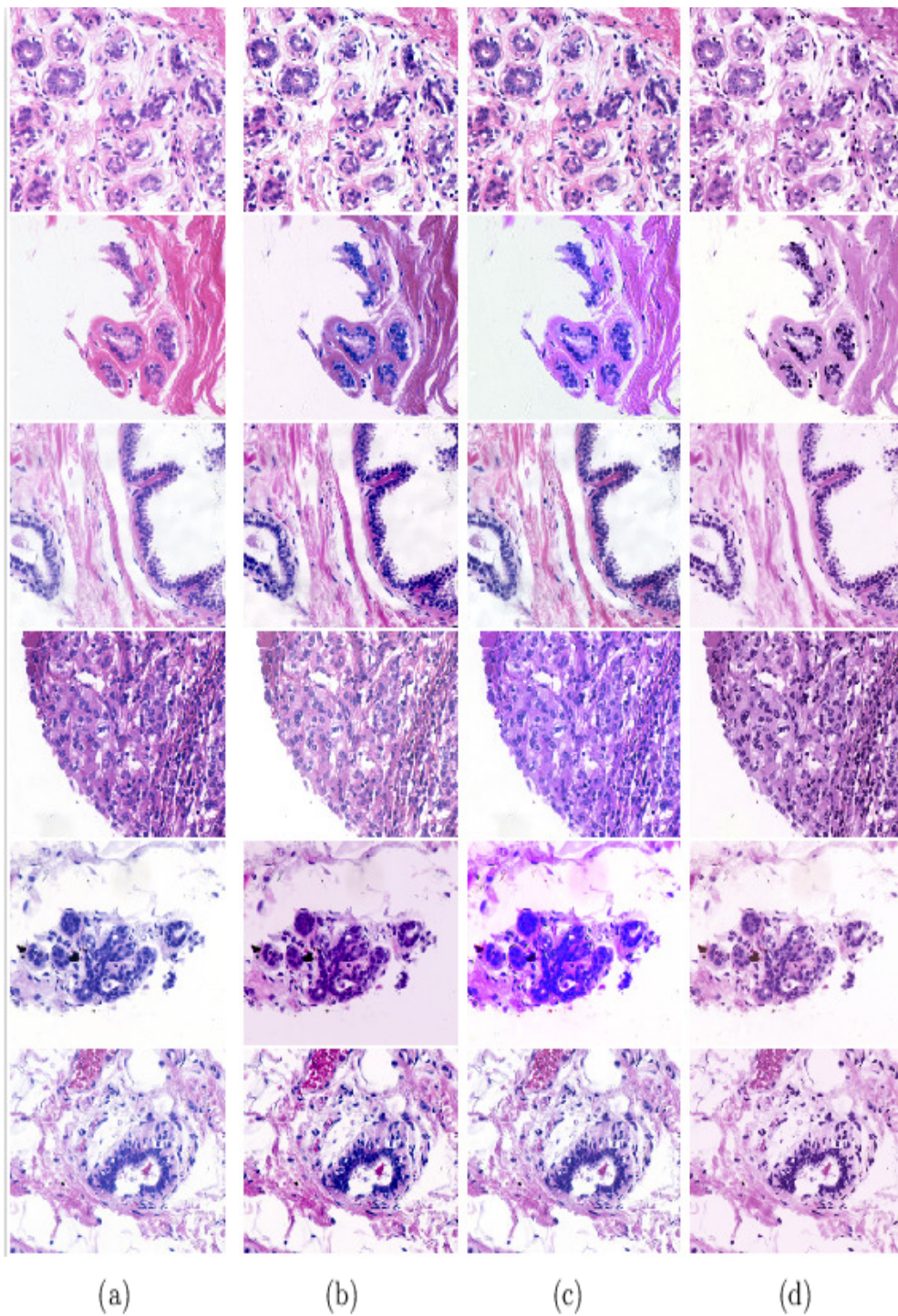
Figure 3.3: Original and color normalized images of UCSB data set obtained using different methods : (a) Original, (b) CT, (c) PF and (d) SCD.

separation as well as color normalization approaches, in terms of satisfying the physical constraints and maintaining the within-image color consistency after color normalization.
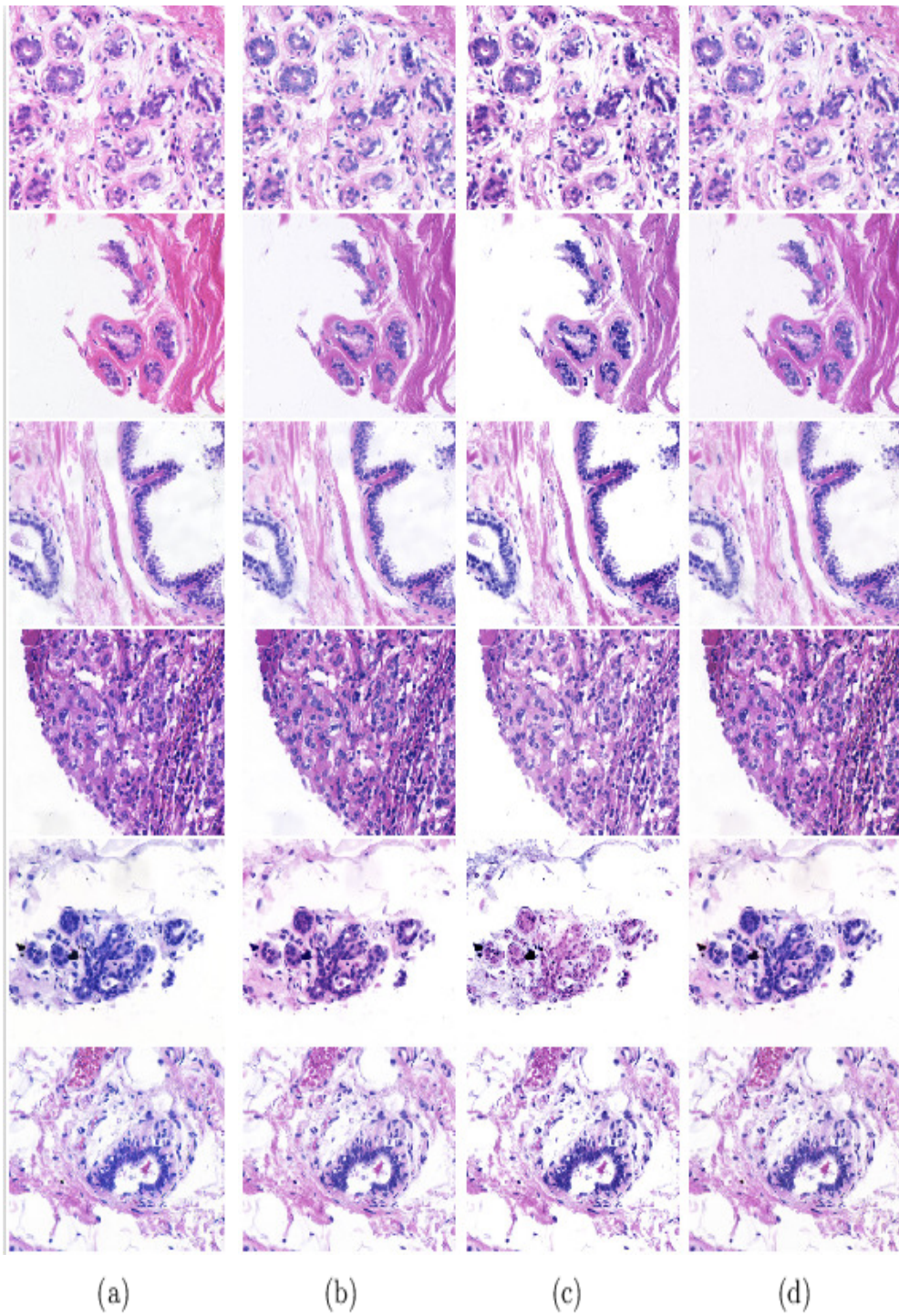
Figure 3.4: Original and color normalized images of UCSB data set obtained using different methods : (a) Original, (b) HTN, (c) SPCN and (d) RFCC$_{\text{cosine}}$.

Moreover, the proposed method does not need any prior information other than the information about the number of stains involved in the staining treatment of histological images

under consideration.

The cosine distance used in this chapter is defined based on the cosine function. However, the cosine function cannot model concentration of circular hue values near the mean of each stain class and fails to capture the intrinsic data distribution properly. The area under the curve of cosine function can be varied, without affecting its periodicity, by multiplying any positive parameter. Based on this concept, the next chapter introduces a new dissimilarity measure to deal with the circular nature of hue values.

Chapter 4

A New Dissimilarity Measure for Circular Clustering in Fuzzy Approximation Spaces

4.1 Introduction

In histology, microscopic images of tissue sections are examined to study the manifestation of diseases under consideration. The most important property of histological images is the enormous density of data compared to other imaging modalities, which makes computer-aided diagnosis more accurate than other modalities. To facilitate pathologists' examination, tissue samples are stained with multiple contrasting histochemical reagents, which in turn highlight different tissue structures and cellular features [42]. Hence, color in pathology plays a pivotal role as a good indicator of histological components.

One of the most common and primary problems of histological tissue analysis is the inadmissible inter and intra-specimen variation in stained tissue color. There are a number of factors associated with the inconsistency in color representation, which has been discussed in Chapter 3. Consequently, numerical features extracted from histological images may lead to difficulty in image interpretation by automated systems, trained on a specific stain color appearance [15, 76]. Although some methods have been proposed to maintain color consistency in images that follow Lambertian (reflective) model of image formation [44], these methods become irrelevant for histological images formed via transmitted light microscopy.

One of the simplest approaches of color normalization is histogram specification [20]. Recently, variations of histogram specification have been proposed in [116] and [4] based on color map quantile matching and histogram landmark matching, respectively. Since histogram matching based approaches ignore actual causes of color disagreement, significant amount of histological information is lost after color normalization. In [38], each image is transformed from correlated RGB to decorrelated $L\alpha\beta$ color space [34], and query image channels are standardized based on the template image channel statistics. In [144], the images are first converted from RGB to HSI color space, and based on the angular differences,

the query image channels are normalized using template image channels. The approaches reported in [38,144] assume that the proportions of tissue compartments are identical across the images being normalized, which is not a legitimate assumption in histological image analysis.

The stain separation based color normalization methods try to address the above problems to a great extent. Based on the relationship between stain amount and light absorption, as given by the Beer-Lambert law of colorimetry [128], the color deconvolution method is used in [7,8] to extract light absorbing stains. Accordingly, stain color appearance map and corresponding stain density map are estimated using the control slides. For estimating image-specific stain concentration map, global stain color descriptor and local pixel-level RGB information are used in [15] as feature map, while relevance vector machine is used as supervised color classifier. Based on the color linear unmixing model of fluorescence imaging, which corresponds to the blind source separation problem, several approaches have been proposed [19,59]. The modified PF approach [86] uses prior knowledge on color vector direction for the estimation of accurate stain vectors. The major drawback of the plain fitting based approaches is that the associated thresholds and control parameters are prefixed and cannot be updated adaptively according to image content, which may lead to faulty stain decomposition.

In a blind color decomposition method [81], input image is first transformed from RGB color space to Maxwellian chromaticity plane for the identification of reference colors from image information. However, the drawback of this approach is that it ignores the presence of weak stains in the image and is likely to be affected by achromatic portions while estimating the weak stain spectra. In [132], saturation-weighted (SW) hue histogram is computed from each image and k-means is employed to perform pixel clustering. In [133], circular thresholding [139] was applied on the SW hue histogram to extract cluster information, and respective stain matrix was estimated. As these methods ignore color mixing due to stain overlap and hardly assign each pixel to a particular stain cluster, the resultant stain decomposition is supposed to be less accurate. Recently, deep-learning approaches, particularly based on adversarial networks [6,41,90,125], have become popular for stain color normalization of histological images. Under this framework, a model is able to learn image-specific color transformation, data-specific staining properties and task-specific networks [6].

As discussed in Chapter 3, one of the main problems in histological image analysis is uncertainty. Some of the sources of this uncertainty include incompleteness and vagueness in stain class definition, as well as overlapping characteristics of histochemical stains. In [134], Li *et al.* proposed a circular mixture modeling based stain decomposition method, which considers physical constraints such as non-negativity of the factor matrices, and the overlapping nature of associated stains. It assumes that each stain class follows circular normal or von Mises distribution. This distribution attains highest probability density value only at its mean. So, a single hue value, which is equal to the mean, ensures the definite belongingness of a pixel to the stain class, as it can only possess highest probability of belongingness to the stain class. However, this normality assumption induces certain limitations as a single hue value is designated to represent each stain class. The probability of belongingness for a pixel to the stain class will be highest if it attains any of the hue values within that range, and will decrease if it deviates further from the bound. Hence, instead of using only von Mises distribution to represent a stain class for histological images, each

stain class should be defined using two sets: a core region, consisting of multiple hue values to represent definite belongingness of pixels, and an overlapping boundary region. In this background, rough sets [147] and fuzzy set [71], together may be effective to introduce the above concept while handling uncertainty present in histological image analysis. Since the integrated approach has the capability of providing a stronger paradigm for uncertainty handling, it has greater promise in image analysis, where fuzzy set and/or rough sets [104] are being effectively used and proved to be successful.

In this regard, this chapter presents a new circular clustering algorithm, termed as rough-fuzzy circular clustering, for color normalization of histological images. The proposed algorithm, similar to Chapter 3, assumes that each stain class is represented by a centroid, a crisp lower approximation, and a fuzzy boundary. While the membership function of fuzzy set enables efficient handling of overlapping stain classes, the concept of lower approximation and boundary region of rough sets deals with uncertainty, vagueness, and incompleteness in stain class definition. Integrating judiciously the merits of both saturation-weighted hue histogram and local neighborhood information of hue image, the proposed algorithm works on a weighted hue histogram in HSI color space. A new dissimilarity measure is introduced to deal with the circular nature of hue values. The efficacy of the proposed algorithm, along with a comparison with existing stain separation and color normalization algorithms, is demonstrated on benchmark histological image sets. Some of the results of this chapter are reported in [105].

The organization of the rest of this chapter is as follows: Section 4.2 presents a new rough-fuzzy circular clustering algorithm based on the proposed new dissimilarity measure, along with the estimation of associated parameters and other theoretical analysis, to address the stain color normalization problem of histological images. The efficacy of the proposed method is analyzed with respect to several state-of-the-art stain separation and color normalization approaches on publicly available histological image data sets in Section 4.3. Concluding remarks are provided in Section 4.4.

4.2 Proposed Stain Color Normalization Method

Stain decomposition is a process that estimates the stain color appearance map and density proportions at each pixel (stain density map) in a histological image. In the proposed method, the stain decomposition problem is formulated following transmission light microscopy [132] based image model. Geometrically, as the stain vectors intersect at the origin in the optical density (OD) space, an image is first converted to OD domain and then factorized to obtain the stain color appearance map and corresponding stain density map. Fig. 4.1 depicts different components of the proposed method. To estimate the stain color appearance map, a rough-fuzzy circular clustering algorithm is proposed in fuzzy approximation spaces. It is based on a new dissimilarity measure that follows next.

4.2.1 Circular Dissimilarity Measure

The color distribution of the image pixels in hue channel of the HSI color space is studied to achieve robustness. It is based on the intuition that hue is a natural descriptor of the color distribution in a color image. As in HSI space, chroma information (hue and saturation)
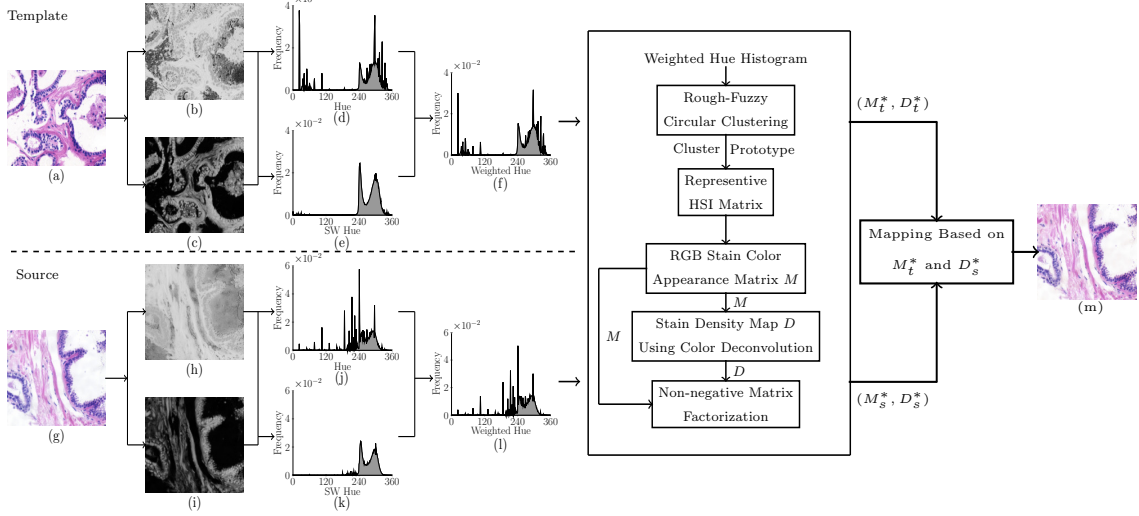
Figure 4.1: Proposed method for stain color normalization of histological images: (a) and (g) represent the template and source images, respectively. (b) and (h) depict the corresponding hue information, while (c) and (i) are the respective saturation images. The hue histograms of template and source images are given in (d) and (j), while saturation-weighted (SW) hue histograms are depicted in (e) and (k), respectively. (f) and (l) are the weighted hue histograms of template and source images, respectively, obtained from corresponding hue and SW hue histograms. The proposed rough-fuzzy circular clustering algorithm accepts weighted hue histograms of (f) and (l) as inputs and generates cluster prototypes as corresponding outputs. The cluster prototypes of both template and source images are used to generate corresponding HSI matrices. The stain color appearance matrix M and stain density map D are estimated from HSI matrix, where non-negative matrix factorization ensures non-negativity constraint of both M and D. Based on M and D, the color normalized source image, corresponding to (g), is produced and reported in (m).

is separated from the luma information (illumination), color analysis in the hue domain is invariant to irregular illumination. So, each histological image is first converted from RGB to HSI color space for further analysis. In this section, a new dissimilarity measure is introduced to deal with the circular nature of hue values.

Let θ_i and θ_j denote the i-th and j-th hue values, respectively, and $\theta_0 = (\theta_i - \theta_j)$ represents the difference between these two hue values. In general, the similarity or dissimilarity between θ_i and θ_j can be computed by either $\cos(\theta_0)$ or $(1 - \cos(\theta_0))$. When two hue values θ_i and θ_j coincide, the angular deviation θ_0 becomes 0 and eventually, $\cos(\theta_0)$ attains its maximum value. The variation of $\cos(\theta)$ with respect to the angular value θ is presented in Fig. 4.2(a), while Fig. 4.2(f) depicts that of $(1 - \cos(\theta))$. So, the similarity between two hue values can be captured by the cosine function.

However, the cosine function fails to model concentration of values near its peak. The area under the curve of cosine function can be varied, without affecting its periodicity, by multiplying a parameter $\kappa > 0$ with $\cos(\theta_0)$. The variation of $\kappa \cos(\theta)$ with respect to θ is presented in Fig. 4.2(b), for different values of κ. If we consider the whole angular range $[0, 2\pi)$, the definite integral $\int_0^{2\pi} \kappa \cos(\theta) d\theta = 0$. So, to incorporate the effect of κ in defining
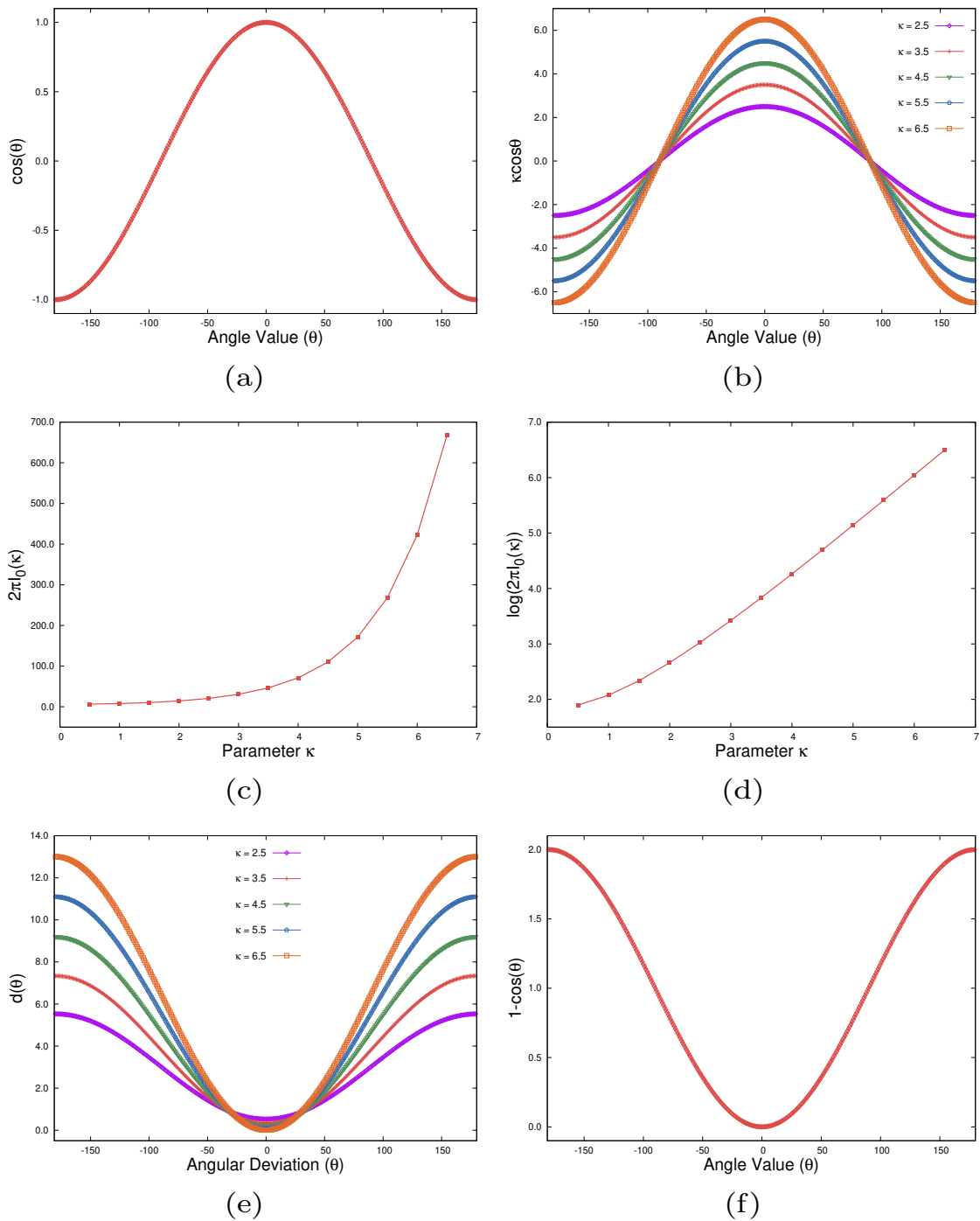
Figure 4.2: Behavior of different expressions involved in the development of proposed dissimilarity measure: (a) $\cos(\theta)$ versus angle value θ, (b) $\kappa \cos(\theta)$ versus angle value θ, (c) $2\pi I_0(\kappa)$ versus concentration parameter κ, (d) $\log(2\pi I_0(\kappa))$ versus concentration parameter κ, (e) proposed dissimilarity measure $d(\theta)$ versus angular deviation θ and (f) cosine distance measure $(1 - \cos(\theta))$ versus angular deviation θ.

the dissimilarity measure, the following exponential term can be considered:

$$\int_0^{2\pi} \exp(\kappa \cos(\theta))d\theta$$

as exponential of any finite entity is always positive. This term is equal to $2\pi I_0(\kappa)$, where $I_0(.)$ is the modified Bessel function of first kind and order zero, defined as:

$$I_0(\kappa) = \frac{1}{2\pi} \int_0^{2\pi} \exp(\kappa \cos(\theta))d\theta. \tag{4.1}$$

Fig. 4.2(c) presents the variation of $2\pi I_0(\kappa)$ as a function of κ, which grows exponentially as the value of parameter κ increases. Fig. 4.2(d) presents the variation of $\log(2\pi I_0(\kappa))$ as a function of κ, which grows almost linearly as the value of parameter κ increases. This term is also constant for a particular value of κ. Based on the above concept, a new dissimilarity measure, denoted by $d(\theta_i, \theta_j)$, is defined as follows:

$$\begin{aligned}
d(\theta_i, \theta_j) &= \log\left(\int_0^{2\pi} \exp(\kappa \cos(\theta))d\theta\right) - \kappa \cos(\theta_i - \theta_j); \\
&= \log(2\pi I_0(\kappa)) - \kappa \cos(\theta_i - \theta_j); \tag{4.2}
\end{aligned}$$

The nature of the proposed dissimilarity measure $d(\theta)$ with respect to θ is reported in Fig. 4.2(e), for different values of κ. The behavior of standard circular distance $(1 - \cos(\theta))$ with respect to θ is reported in Fig. 4.2(f). It can be noticed from Fig. 4.2(e) and Fig. 4.2(f) that the proposed dissimilarity measure is effective in modeling the concentration of hue values around the mean by varying the parameter κ, which is not possible by the existing cosine dissimilarity measure.

From (4.2), the following properties can be derived for the proposed dissimilarity measure:

1. $d(\theta_i, \theta_j) = d(\theta_j, \theta_i)$;

2. $d(\theta_i, \theta_i) < d(\theta_i, \theta_j), \forall j \neq i$;

3. $d(\theta_i, \theta_k) \leqslant d(\theta_i, \theta_j) + d(\theta_j, \theta_k), \forall \theta_i, \theta_j, \theta_k \in [0, 2\pi)$.

The first two properties are trivial. The first one ensures the symmetric property of the proposed dissimilarity measure. The second one states that whenever θ_i and θ_j coincide, the dissimilarity measure is the minimum. The third axiom, the triangle inequality, seems to be intuitively evident, but this property is a hard one to satisfy.

4.2.2 Weighted Hue Histogram

The proposed rough-fuzzy circular clustering algorithm considers a weighted hue histogram H, instead of standard hue histogram, to estimate the stain class representatives. The weighted hue histogram H judiciously integrates the merits of both saturation-weighted hue histogram H^{sw} and local neighborhood information. In [10], it has been shown that the standard hue histogram consists of a number of sharp ridges attributed by achromatic

pixels. So, the hue values, corresponding to achromatic region, are ill-defined and insignificant for the pixels that contain small saturation values. As a result, analysis of color distribution on a standard hue histogram, corrupted by these insignificant pixels, yields insignificant and faulty output. So, the effects of these achromatic pixels on the standard hue histogram can be reduced by computing saturation-weighted hue histogram H^{sw} [10], as follows:

$$H^{sw}(\theta) = \sum_{k \in \mathcal{I}} s_k \delta(\theta, h_k); \tag{4.3}$$

$$\text{where} \quad \delta(\theta, h_k) = \begin{cases} 1 & \text{if } \theta = h_k \\ 0 & \text{otherwise.} \end{cases} \tag{4.4}$$

Here, $\mathcal{I}$ represents a pathology image, s_k and h_k denote, respectively, saturation and hue value corresponding to the k-th pixel in the HSI color space. In order to consider the local neighborhood information of a pixel in standard hue image h, the proposed algorithm introduces a α-controlled hue image, following the model of Szilagyi $et\ al.$ [78]. The proposed α-controlled mean filtered hue image ξ can be obtained from the original hue image h as follows:

$$\xi_k = \frac{1}{1 + \alpha} \left(h_k + \frac{\alpha}{|N_k|} \sum_{h_j \in N_k} h_j \right); \tag{4.5}$$

where ξ_k denotes the k-th pixel of the α-controlled hue image ξ and N_k represents the set of neighbors within a window around h_k. The parameter α controls the influence of the neighborhood information. A α-controlled approximated histogram H^{α} can be computed from the α-controlled hue image ξ, where $H^{\alpha}(\theta)$ denotes the number of pixels in ξ having hue value θ. Combining both saturation-weighted hue histogram H^{sw} and α-controlled approximated histogram H^{α}, a weighted hue histogram H is defined as follows:

$$H(\theta) = \frac{1}{2} \left[H^{sw}(\theta) + H^{\alpha}(\theta) \right]. \tag{4.6}$$

4.2.3 Rough-Fuzzy Circular Clustering

The proposed rough-fuzzy circular clustering algorithm assumes that each stain class is represented by a centroid, a crisp lower approximation, and a fuzzy boundary. Let $\underline{A}(\beta_i)$ and $B(\beta_i)$ denote the lower approximation and boundary region of the i-th stain class β_i. The upper approximation $\overline{A}(\beta_i)$ of stain class β_i is given by $\overline{A}(\beta_i) = [\underline{A}(\beta_i) \cup B(\beta_i)]$. According to the definitions of lower approximation and boundary region of rough sets [147], if the j-th hue value θ_j belongs to the lower approximation $\underline{A}(\beta_i)$ of the i-th stain class β_i, then θ_j does not belong to lower approximations and boundary regions of any other stain classes. That is, θ_j definitely belongs to the stain class β_i. Hence, the weights of hue values lying in lower approximation of a stain class should be independent of other centroids and stain classes, and should not be coupled with their similarity with respect to other centroids. On the other hand, if a hue value θ_j belongs to the boundary region $B(\beta_i)$ of the i-th stain class β_i, then θ_j possibly belongs to β_i and potentially belongs to another stain class.

The proposed rough-fuzzy circular clustering algorithm with parameter set ψ divides the weighted hue histogram H, corresponding to a histology image $\mathcal{I}$, into c stain classes

by minimizing the following objective function:

$$J_{\mathrm{RF}}(\psi) = \sum_{i=1}^{c} \left[\omega \times J_i^L(\psi) + (1 - \omega) \times J_i^B(\psi) \right]; \qquad (4.7)$$

where ω regulates the relative importance between lower approximation and boundary region of each stain class, and

$$J_i^L(\psi) = \sum_{\theta_j \in \underline{A}(\beta_i)} [\log(2\pi I_0(\kappa_i)) - \kappa_i \cos(\theta_j - \mu_i)] H(\theta_j);$$

$$J_i^B(\psi) = \sum_{\theta_j \in B(\beta_i)} \nu_{ij}^m [\log(2\pi I_0(\kappa_i)) - \kappa_i \cos(\theta_j - \mu_i)] H(\theta_j) + \sum_{\theta_j \in B(\beta_i)} [\nu_{ij}^m \log(\nu_{ij}^m) - \nu_{ij}^m] H(\theta_j).$$

Here θ_j and μ_i represent the j-th hue and centroid of the i-th stain class β_i, ν_{ij} denotes fuzzy membership of the j-th hue into i-th class β_i, $m \in [1, \infty)$ is the fuzzifier, κ_i is the concentration parameter corresponding to β_i, and $I_0(\cdot)$ is the modified Bessel function of first kind and order zero, defined in (4.1). So, the parameter set ψ is given as $\psi = \{\mu_i, \kappa_i, \nu_{ij}\}$. Note that the term $[\nu_{ij}^m \log(\nu_{ij}^m) - \nu_{ij}^m]$ is a monotonically decreasing function of ν_{ij} in $[0, 1]$, and forces ν_{ij} to be as large as possible to avoid the trivial solution.

4.2.3.1 Estimation of Membership Function

To estimate the fuzzy membership ν_{ij} of the j-th hue into i-th stain class β_i, (5.2.3) is partially differentiated with respect to ν_{ij} as follows:

$$\frac{\partial J_{\mathrm{RF}}(\psi)}{\partial \nu_{ij}} = 0;$$

$$\Rightarrow m\nu_{ij}^{m-1}[\log(2\pi I_0(\kappa_i)) - \kappa_i \cos(\theta_j - \mu_i)] H(\theta_j) + m\nu_{ij}^{m-1} \log(\nu_{ij}^m) H(\theta_j) = 0;$$

$$\Rightarrow \log(\nu_{ij}^m) = \kappa_i \cos(\theta_j - \mu_i) - \log(2\pi I_0(\kappa_i));$$

$$\Rightarrow \log(\nu_{ij}^m) = \log\left[\frac{\exp\{\kappa_i \cos(\theta_j - \mu_i)\}}{2\pi I_0(\kappa_i)}\right];$$

$$\Rightarrow \nu_{ij} = \left[\frac{\exp\{\kappa_i \cos(\theta_j - \mu_i)\}}{2\pi I_0(\kappa_i)}\right]^{\frac{1}{m}}. \qquad (4.8)$$

From (4.8), it is evident that the fuzzy membership function follows von Mises distribution, which is a variant of the Gaussian distribution in circular domain and is defined next.

Definition 4.1. *A circular random variable θ is said to follow von Mises or circular normal distribution, if it has the probability density function defined as follows [96]:*

$$vM(\mu, \kappa) = \frac{1}{2\pi I_0(\kappa)} \exp\{\kappa \cos(\theta - \mu)\}; \qquad (4.9)$$

where $\mu \in [0, 2\pi)$ *and* $\kappa > 0$ *denote the mean direction and concentration parameter, respectively. Fig. 4.3 presents the von Mises distribution for different values of concentration parameter* κ.
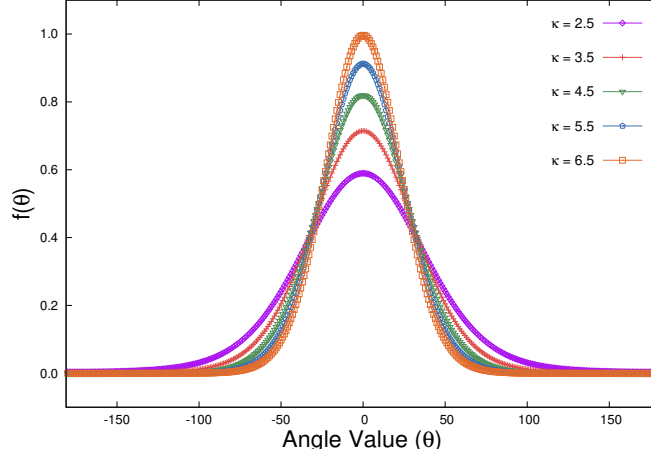
Figure 4.3: von Mises distribution for different values of parameter κ.

4.2.3.2 Computation of Cluster Prototype

The new centroid is computed based on the weighted average of the crisp lower approximation and fuzzy boundary region of each stain class. The centroid, corresponding to each stain class β_i, is computed by solving (4.7) with respect to μ_i, as follows:

$$\frac{\partial J_{\mathrm{RF}}(\psi)}{\partial \mu_i} = 0;$$

$$\Rightarrow \omega \times \sum_{\theta_j \in \underline{A}(\beta_i)} \sin(\theta_j - \mu_i) H(\theta_j) + (1 - \omega) \times \sum_{\theta_j \in B(\beta_i)} \nu_{ij}^m \sin(\theta_j - \mu_i) H(\theta_j) = 0;$$

$$\Rightarrow \mu_i = \arctan\left[\frac{\omega \times \mathcal{A}_1 + (1 - \omega) \times \mathcal{A}_2}{\omega \times \mathcal{B}_1 + (1 - \omega) \times \mathcal{B}_2} \right]; \tag{4.10}$$

$$\text{where} \quad \mathcal{A}_1 = \sum_{\theta_j \in \underline{A}(\beta_i)} \sin(\theta_j) H(\theta_j); \quad \mathcal{A}_2 = \sum_{\theta_j \in B(\beta_i)} \nu_{ij}^m \sin(\theta_j) H(\theta_j);$$

$$\mathcal{B}_1 = \sum_{\theta_j \in \underline{A}(\beta_i)} \cos(\theta_j) H(\theta_j); \quad \text{and} \quad \mathcal{B}_2 = \sum_{\theta_j \in B(\beta_i)} \nu_{ij}^m \cos(\theta_j) H(\theta_j).$$

From (4.10), it is observed that the computation of centroid is dependent on the choice of the parameter ω. Since the hue values lying in lower approximation region definitely belong to a certain stain class, they are supposed to be assigned a higher weightage ω compared to $(1 - \omega)$ corresponding to the hues lying in boundary region.

4.2.3.3 Estimation of Concentration Parameter

To estimate the concentration parameter κ_i, corresponding to the i-th stain class β_i, (4.7) is partially differentiated with respect to κ_i, as follows:

$$\frac{\partial J_{\mathrm{RF}}(\psi)}{\partial \kappa_i} = 0;$$

$$\Rightarrow \omega \times \sum_{\theta_j \in \underline{A}(\beta_i)} \left\{ \frac{I_0'(\kappa_i)}{I_0(\kappa_i)} - \cos(\theta_j - \mu_i) \right\} H(\theta_j) +$$

$$(1 - \omega) \times \sum_{\theta_j \in B(\beta_i)} \nu_{ij}^m \left\{ \frac{I_0'(\kappa_i)}{I_0(\kappa_i)} - \cos(\theta_j - \mu_i) \right\} H(\theta_j) = 0;$$

$$\Rightarrow \frac{I_1(\kappa_i)}{I_0(\kappa_i)} \left\{ \omega \times \sum_{\theta_j \in \underline{A}(\beta_i)} H(\theta_j) + (1 - \omega) \times \sum_{\theta_j \in B(\beta_i)} \nu_{ij}^m H(\theta_j) \right\}$$

$$= \left\{ \omega \times \sum_{\theta_j \in \underline{A}(\beta_i)} \cos(\theta_j - \mu_i) H(\theta_j) + (1 - \omega) \times \sum_{\theta_j \in B(\beta_i)} \nu_{ij}^m \cos(\theta_j - \mu_i) H(\theta_j) \right\}$$

as $I_0'(\kappa_i) = \frac{d}{d\kappa_i}(I_0(\kappa_i)) = I_1(\kappa_i)$. Let $\mathsf{T}(\kappa_i) = \frac{I_1(\kappa_i)}{I_0(\kappa_i)}$. So,

$$\kappa_i = \mathsf{T}^{-1} \left[\frac{\omega \times \mathcal{X}_1 + (1 - \omega) \times \mathcal{X}_2}{\omega \times \mathcal{Y}_1 + (1 - \omega) \times \mathcal{Y}_2} \right]; \tag{4.11}$$

$$\text{where} \quad \mathcal{X}_1 = \sum_{\theta_j \in \underline{A}(\beta_i)} \cos(\theta_j - \mu_i) H(\theta_j); \quad \mathcal{X}_2 = \sum_{\theta_j \in B(\beta_i)} \nu_{ij}^m \cos(\theta_j - \mu_i) H(\theta_j);$$

$$\mathcal{Y}_1 = \sum_{\theta_j \in \underline{A}(\beta_i)} H(\theta_j); \quad \text{and} \quad \mathcal{Y}_2 = \sum_{\theta_j \in B(\beta_i)} \nu_{ij}^m H(\theta_j).$$

Here, $\mathsf{T}^{-1}(.)$ is approximated using numerical methods [96].

Each stain class β_i is characterized by the parameter set $\psi = \{\mu_i, \kappa_i, \nu_{ij}\}$. The cluster prototype μ_i is considered as the representative hue for the i-th stain class β_i. The parameter κ_i controls the concentration of hue values around the centroid μ_i of the stain class β_i. The fuzzy membership ν_{ij} of the j-th hue into i-th stain class β_i depends on both μ_i and κ_i.

4.2.4 Proposed Algorithm and Stain Color Normalization

Approximate optimization of $J_{\mathrm{RF}}(\psi)$ in (4.7) by the proposed rough-fuzzy circular clustering algorithm is performed based on Picard iteration through (4.8), (4.10) and (4.11). The process starts by initializing $\mu_i^{(0)}$, $\kappa_i^{(0)}$, and $\nu_{ij}^{(0)}$ for $(c - 1)$ stain regions and one achromatic region. As hue is circular in nature, the circular thresholding method [139] is used to generate approximate partitions corresponding to achromatic region and different stain

50

regions. The initial cluster prototype $\mu_i^{(0)}$ and concentration parameter $\kappa_i^{(0)}$, corresponding to both achromatic and stain regions, are computed from the weighted hue histogram H, based on the initial parameter estimation procedure reported in [134]. The fuzzy memberships of all hues are then calculated using (4.8). After computing ν_{ij} for c stain classes (including achromatic region) and n hue values, the values of ν_{ij} for each hue θ_j are sorted and the difference of two highest memberships of θ_j is compared with a threshold value λ. Let ν_{ij} and ν_{kj} be the highest and second highest memberships of θ_j. If $(\nu_{ij} - \nu_{kj}) > \lambda$, then $\theta_j \in \underline{A}(\beta_i)$, otherwise $\theta_j \in B(\beta_i)$ and $\theta_j \in B(\beta_k)$. After assigning all hues in lower approximations or boundary regions of different stain classes based on λ, memberships ν_{ij} of the hues are modified. The values of ν_{ij} are set to 1 for the hues belonging to lower approximations, while those in boundary regions remain unchanged. The new centroids of the stain classes are calculated as per (4.10), while corresponding concentration parameters are computed using (4.11). The main steps of proposed rough-fuzzy circular clustering algorithm are as follows:

Algorithm 4.1 Proposed rough-fuzzy circular clustering algorithm.

Input: Weighted hue histogram H, number of stain classes c.
Output: Estimated parameter set $\psi = \{\mu_i, \kappa_i, \nu_{ij}\}$.

1: Initialize $\mu_i^{(0)}$ and $\kappa_i^{(0)}$ $(1 \leqslant i \leqslant c-1)$ for $(c-1)$ stain regions and $\mu_{ac}^{(0)}$ and $\kappa_{ac}^{(0)}$ for achromatic region.
2: Choose values for fuzzifier m, threshold λ and tolerance value ϵ. Set iteration counter $t = 1$.
3: Compute ν_{ij} for c stain classes and n hues using (4.8).
4: If ν_{ij} and ν_{kj} be the two highest memberships of θ_j and $(\nu_{ij} - \nu_{kj}) > \lambda$, then $\theta_j \in \underline{A}(\beta_i)$; otherwise $\theta_j \in B(\beta_i)$ and $\theta_j \in B(\beta_k)$.
5: Modify ν_{ij} considering lower approximation and boundary regions for c stain classes and n hues.
6: Compute new centroid μ_i as per (4.10).
7: Compute concentration parameter κ_i using (4.11).
8: Repeat steps 3 to 7 until $|\nu_{ij}^{(t)} - \nu_{ij}^{(t-1)}| < \epsilon$.
9: Stop.

The proposed circular clustering algorithm partitions the hue values into two regions, namely, lower approximation and boundary region, based on the value of λ. In practice, the following definition of λ works well:

$$\lambda = \frac{1}{n} \sum_{j=1}^{n} (\nu_{ij}^{(0)} - \nu_{kj}^{(0)}); \tag{4.12}$$

where n is the total number of hues, $\nu_{ij}^{(0)}$ and $\nu_{kj}^{(0)}$ are the highest and second highest memberships of θ_j at $t = 0$. So, λ represents the average difference of two highest memberships of all hue values present in an image. A good clustering procedure should make the value of λ as high as possible.

The stain color appearance map can be computed from the centroid of each stain class. The cluster prototype μ_i, $i = 1, 2, \cdots, c$, estimated by the proposed rough-fuzzy circu-

lar clustering, is considered as the representative hue h_i for the i-th stain class β_i. The saturation-weighted statistics, described in [132], is used to compute the corresponding saturation s_i and intensity v_i for obtaining the stain representative color. The representative color $I_i^{\text{hsi}} = [h_i, s_i, v_i]^T$, corresponding to the i-th stain class, in HSI domain is first converted to RGB color space: $I_i^{\text{rgb}} = F_{\text{hsi}}^{\text{rgb}}(I_i^{\text{hsi}})$, where $F_{\text{hsi}}^{\text{rgb}}(.)$ denotes a function for converting a stain vector from HSI domain to RGB domain. The absorption vector corresponding to I_i^{rgb} in OD domain is given as $M_i^{\text{rgb}} = \log(I^b/I_i^{\text{rgb}})$, where I^b represents the intensity value corresponding to a background pixel. The stain density value, associated with pixel p, can be computed using the following relation:

$$D^{\text{rgb}}(p) = (M^{\text{rgb}})^{-1}[\log(I^b/I(p))]. \tag{4.13}$$

However, in OD domain, the image colors are represented as the non-negative combination of c stain vectors associated with the staining routine. Due to the non-negativity constraint of both stain color appearance matrix M and stain density map D, NMF is utilized to refine candidate stain decomposition solution by minimizing the following decomposition error:

$$\{\acute{M}, \acute{D}\} = \arg \min_{M \geq 0, D \geq 0} \| \log(I^b/I) - MD \|^2; \tag{4.14}$$

where ≥ 0 denotes element-wise non-negativity of the matrix.

4.2.5 Convergence Analysis

In this section, a mathematical analysis on the convergence condition of the proposed rough-fuzzy circular clustering algorithm is presented. To examine the convergence property of the proposed algorithm, the Hessian (second order partial derivative) matrices with respect to the parameters involved in the objective function are computed and then checked whether the Hessian matrices are positive definite or not. The second order partial derivative of the objective function in (4.7) with respect to the cluster centroid μ_i leads to

$$\frac{\partial^2 J_{\text{RF}}(\psi)}{\partial \mu_i^2} = \omega \times U_i + (1 - \omega) \times V_i; \tag{4.15}$$

$$\text{where} \quad U_i = \sum_{\theta_j \in \underline{A}(\beta_i)} \kappa_i \cos(\theta_j - \mu_i) H(\theta_j); \tag{4.16}$$

$$\text{and} \quad V_i = \sum_{\theta_j \in B(\beta_i)} \kappa_i \nu_{ij}^m \cos(\theta_j - \mu_i) H(\theta_j). \tag{4.17}$$

It is evident that $\kappa_i > 0 \; \forall i$, $\nu_{ij}^m \geq 0$, $\cos(\theta_j - \mu_i) \geq 0$ and $H(\theta_j) \geq 0 \; \forall \theta_j$. Hence, for each cluster β_i, $U_i > 0$ and $V_i > 0$. Also, it can be easily verified that, $\forall k \neq i$,

$$\frac{\partial^2 J_{\text{RF}}(\psi)}{\partial \mu_k \partial \mu_i} = 0. \tag{4.18}$$

Hence, it is clear from (4.18) that all the off-diagonal entries in the respective Hessian matrices are 0. The proposed algorithm is guaranteed to converge if the Hessian matrix with respect to μ_i is diagonally dominant [2]. This is a sufficient condition, not a necessary one. If the Hessian matrix is not diagonally dominant, the iteration may or may not converge. The Hessian matrix corresponding to (4.16) is given by $\mathcal{P} = \text{diag}\{U_1, \cdots, U_i, \cdots, U_c\}$ and the matrix corresponding to (4.17) is given by $\mathcal{Q} = \text{diag}\{V_1, \cdots, V_i, \cdots, V_c\}$.

It can be easily observed that the Hessian matrices $\mathcal{P}$ and $\mathcal{Q}$ are diagonally dominant. So, $\mathcal{P}$ and $\mathcal{Q}$ are certainly positive definite matrices. Hence, all the eigenvalues corresponding to $\mathcal{P}$ and $\mathcal{Q}$ are positive. Under this condition, the iterative algorithm would converge to at least a local optimum solution if (4.10) was repetitively applied keeping both ν_{ij} and κ_i constant. Similarly, it can be easily shown that the Hessian matrices with respect to both ν_{ij} and κ_i are diagonal positive definite matrices. Intuitively, the objective function $J_{\text{RF}}(\psi)$ reduces in all steps corresponding to (4.8), (4.10) and (4.11), which makes the system strictly descent. It ensures the convergence of the proposed rough-fuzzy circular clustering algorithm.

4.3 Performance Analysis

The performance of the proposed rough-fuzzy circular clustering based stain separation and color normalization method is studied extensively and compared with that of

- several state-of-the-art stain separation methods, namely, plane fitting (PF) [84], enhanced plane fitting (EPF) [89], structure-preserving color normalization (SPCN) [22], expectation-maximization (EM) algorithm [134], and the method named HTN due to Li and Plataniotis [132]; and

- different color normalization algorithms such as color transfer technique (ColTrans) [38], stain color description (SCD) [15], stain normalization using generative adversarial networks (SN-GAN) [41], StainGAN [90], and adversarial stain transfer (AST) [6], as well as PF [84], EPF [89], HTN [132], SPCN [22] and RFCC$_{\text{cosine}}$.

To find out the optimum values of weight parameter ω and control parameter α for the proposed method, the value of ω is varied from 0.51 to 0.99, while that of α is varied in the range of 0.1 to 1.0, with an increment of 0.1. The optimum values of ω and α, for a particular image, are obtained using Silhouette index [103], as discussed in supplementary material.

To evaluate the performance of different stain separation and color normalization methods, following two publicly available histology image sets are used in the current study.

- UCSB Breast Cancer Cell Data: This data set, published by University of California, Santa Barbara [36], comprises of 32 non-cancerous benign cell and 26 cancerous malignant cell images, acquired from ten hematoxylin and eosin (H&E) stained breast cancer biopsy sets. Each image has a resolution of 896×768, and associated ground-truth with nuclei considered as ROIs.

- CMU Data: This data set, published by the bimagicLab in Carnegie Mellon University [89], contains three H&E stained 48-bit linear RGB format images with a

resolution of 1280×1024. In addition, this data set consists of stain decomposition ground-truth, that is, separate H-stained and E-stained images corresponding to each image.

To evaluate the performance of different stain estimation and separation methods, standard deviation, symmetric Kullback-Leibler (KL) divergence and signal-to-noise ratio (SNR) are used, while normalized median intensity (NMI), between-image color constancy (BiCC) and within-set color constancy (WsCC) indices are used to evaluate color consistency of the specific ROI within a single biopsy set after color normalization. A brief description of experimental set-up, definitions of quantitative indices used, and some important results are described in detail in the supplementary material.

4.3.1 Evaluation of Stain Vector Estimation

As the images within the same biopsy set undergo same type of staining routine and similar storage condition, a good spectral estimation algorithm should produce consistent stain vectors for the images of same biopsy set. The proposed circular clustering algorithm judiciously integrates the theory of rough sets, and merits of new circular dissimilarity measure and weighted hue histogram. To establish the importance of new dissimilarity measure over conventional cosine distance for stain vector estimation, extensive experiments are carried out on UCSB data set, and corresponding results are reported in Table 4.1. In Table 4.1, standard deviations (σ) of the estimated stain vectors, corresponding to ten biopsy sets, are summarized, where σ_{H} and σ_{E} represent the standard deviations of the estimated stain vectors of hematoxylin (H) stain and eosin (E) stain, respectively, computed over all images of the same biopsy set. The lower σ value of the respective RGB matrix element ensures the generation of consistent stain vectors for all images of a given biopsy set. The results reported in Table 4.1 confirm that the proposed dissimilarity measure achieves lowest σ values in 19 and 27 cases, out of 30 cases each, for H-stain and E-stain, respectively.

In order to establish the significance of rough-fuzzy clustering over its hard ($\lambda = 0$) and fuzzy ($\lambda = 1$) counter parts, extensive experimentation is carried out on UCSB data set and corresponding results are reported in Table 4.2. From the results reported in Table 4.2, it can be easily observed that the proposed rough-fuzzy clustering algorithm attains lowest values of standard deviation in 18 and 25 cases for H-stain and E-stain, respectively, while hard and fuzzy clustering achieve lowest σ values in 3 and 0, and 9 and 6 cases, respectively.

In order to establish the relevance of weighted hue histogram in stain vector estimation over standard hue, saturation-weighted (SW) hue, and α-controlled hue histograms, extensive experimentation is done on UCSB data set, and corresponding results are reported in Table 4.3. Table 4.3 presents the comparative performance analysis of different hue histograms on UCSB data set. The UCSB data set contains ten different biopsy sets, nine of them contain 6 images each, while another contains 4 images. As the images within each biopsy set are expected to undergo same type of staining routine and similar storage condition, a good spectral estimation algorithm should produce consistent or similar spectrum vectors for the images of same biopsy set. Table 4.3 reports the standard deviation (σ) of the estimated stain vectors of all the images belonging to the same biopsy set. Lower the value of σ, better the stain estimation procedure is. In Table 4.3, standard deviations of the estimated stain vectors over ten biopsy sets are summarized, where the smallest σ value (values are represented in 10^{-2} scale) of respective matrix element is marked in bold

54

Table 4.1: Performance Analysis of Proposed Dissimilarity Measure on UCSB Data Set: σ_X represents the standard deviation of the estimated stain vectors of X-stain, computed over all images of the same biopsy set, and $X \in \{H, E\}$. Each value is represented in 10^{-2} scale. The smallest value of the respective matrix element is marked in bold.

Biopsy Sets	Different Channels	Cosine Distance		Proposed Dissimilarity	
		σ_H	σ_E	σ_H	σ_E
ytma10_010704	Red	3.19	3.03	**2.84**	**2.35**
	Green	1.69	1.07	**1.27**	**0.89**
	Blue	**1.35**	2.30	1.64	**1.86**
ytma12_010804	Red	**3.38**	5.69	3.59	**4.14**
	Green	2.71	2.06	**2.67**	**1.60**
	Blue	**0.96**	1.10	**0.96**	**1.06**
ytma23_022103	Red	1.57	5.36	**1.37**	**4.00**
	Green	1.30	1.79	**1.07**	**1.51**
	Blue	0.67	1.36	**0.58**	**0.77**
ytma49_042003	Red	3.31	**2.53**	**2.66**	2.94
	Green	2.05	**0.55**	**1.72**	0.61
	Blue	0.98	1.07	**0.55**	**0.83**
ytma49_042203	Red	**2.64**	3.28	2.80	**2.77**
	Green	**2.07**	0.91	2.24	**0.58**
	Blue	**0.31**	**1.95**	0.54	2.06
ytma49_042403	Red	**1.78**	3.79	2.22	**2.30**
	Green	**1.07**	1.07	1.16	**0.93**
	Blue	0.66	0.88	**0.49**	**0.87**
ytma49_072303	Red	1.86	3.40	**1.04**	**1.73**
	Green	1.53	1.20	**0.73**	**0.76**
	Blue	0.30	0.83	**0.20**	**0.40**
ytma49_111003	Red	**2.34**	4.65	2.74	**3.60**
	Green	**2.07**	2.19	2.61	**1.58**
	Blue	0.99	1.85	**0.77**	**1.20**
ytma49_111303	Red	1.83	5.85	**1.38**	**3.35**
	Green	1.66	2.47	**1.31**	**1.69**
	Blue	0.92	1.72	**0.59**	**1.17**
ytma55_030603	Red	**2.65**	4.03	2.99	**3.60**
	Green	**2.29**	1.65	2.78	**1.46**
	Blue	1.16	1.76	**0.85**	**1.63**

font. From all the results reported in Table 4.3, it is seen that the proposed weighted hue histogram performs significantly better than other hue histograms on UCSB data set. Out of total 30 cases of E-stain, for ten biopsy sets, the proposed hue histogram attains lowest σ values in 24 cases, while it achieves lowest σ values in 17 cases for H-stain. On the other hand, the SW hue histogram attains lowest σ values in 11 and 0 cases, out of 30 cases each, for H-stain and E-stain, respectively. However, the standard hue histogram achieves lowest σ values in 2 and 4 cases, and α-controlled hue histogram attains in 0 and 2 cases,

Table 4.2: Performance Analysis of Proposed Rough-Fuzzy Circular Clustering Algorithm on UCSB Data Set: σ_X represents the standard deviation of the estimated stain vectors of X-stain, computed over all images of the same biopsy set, and $X \in \{H, E\}$. Each value is represented in 10^{-2} scale. The smallest value of the respective matrix element is marked in bold.

Different Biopsy Sets	Different Channels	Hard		Fuzzy		Rough-Fuzzy	
		σ_H	σ_E	σ_H	σ_E	σ_H	σ_E
ytma10_010704	Red	13.44	4.68	4.05	2.72	**2.84**	**2.35**
	Green	5.54	8.69	2.00	1.03	**1.27**	**0.89**
	Blue	17.29	15.85	**1.47**	1.87	1.64	**1.86**
ytma12_010804	Red	13.24	5.93	**3.45**	4.26	3.59	**4.14**
	Green	5.67	10.92	2.75	1.90	**2.67**	**1.60**
	Blue	21.17	18.49	**0.87**	**0.97**	0.96	1.06
ytma23_022103	Red	1.89	5.21	1.54	4.38	**1.37**	**4.00**
	Green	1.39	1.86	1.09	1.76	**1.07**	**1.51**
	Blue	0.65	1.23	0.65	0.86	**0.58**	**0.77**
ytma49_042003	Red	15.30	5.79	3.13	**2.58**	**2.66**	2.94
	Green	5.02	8.47	1.93	**0.59**	**1.72**	0.61
	Blue	18.37	15.92	0.70	0.96	**0.55**	**0.83**
ytma49_042203	Red	**2.68**	3.24	2.78	3.17	2.80	**2.77**
	Green	**1.85**	0.77	2.08	0.73	2.24	**0.58**
	Blue	0.63	1.99	0.43	**1.84**	**0.54**	2.06
ytma49_042403	Red	2.89	3.38	**2.02**	2.67	2.22	**2.30**
	Green	1.72	1.07	**1.14**	0.96	1.16	**0.93**
	Blue	0.77	1.00	0.70	**0.83**	**0.49**	0.87
ytma49_072303	Red	1.70	3.19	1.09	1.97	**1.04**	**1.73**
	Green	1.29	1.18	0.85	0.84	**0.73**	**0.76**
	Blue	0.36	0.72	0.28	0.63	**0.20**	**0.40**
ytma49_111003	Red	3.05	3.87	2.97	4.58	**2.74**	**3.60**
	Green	2.61	1.85	**2.60**	2.23	2.61	**1.58**
	Blue	1.16	1.59	1.09	1.70	**0.77**	**1.20**
ytma49_111303	Red	1.73	5.44	1.43	3.71	**1.38**	**3.35**
	Green	1.55	2.41	**1.30**	2.03	1.31	**1.69**
	Blue	0.84	1.68	0.88	**1.17**	**0.59**	1.17
ytma55_030603	Red	2.69	4.32	**2.66**	4.61	2.99	**3.60**
	Green	**2.26**	1.89	2.31	1.86	2.78	**1.46**
	Blue	1.04	2.06	**0.77**	1.96	0.85	**1.63**

for H-stain and E-stain, respectively.

Finally, the performance of the proposed approach, in stain vector estimation, is extensively compared with that of several state-of-the-art stain separation methods, namely, PF [84], EPF [89], SPCN [22], EM [134], HTN [132] and RFCC_cosine [104], and corresponding results are reported in Table 4.5. From the results reported in Table 4.5, it is seen that the proposed method attains lowest σ values in 12 and 13 cases, out of 30 cases each, for H-

Table 4.3: Performance Analysis of Different Hue Histograms (Standard Hue Histogram, SW Hue Histogram, α-Controlled Hue Histogram, and Proposed Weighted Hue Histogram) on UCSB Data: σ_X represents the standard deviation of the estimated stain vectors of X-stain, computed over all images of the same biopsy set, and $X \in \{H, E\}$. Each value is represented in 10^{-2} scale. The smallest value of the respective matrix element is marked in bold.

Different Biopsy Sets	Different Channels	Standard		SW		α-controlled		RFCC$_{vM}$	
		σ_H	σ_E	σ_H	σ_E	σ_H	σ_E	σ_H	σ_E
ytma10_010704	Red	10.41	2.86	3.54	4.10	9.55	**1.91**	**2.84**	2.35
	Green	6.71	1.02	2.02	1.08	7.47	0.99	**1.27**	**0.89**
	Blue	6.71	2.49	**1.25**	2.71	1.50	2.56	1.64	**1.86**
ytma12_010804	Red	7.36	5.54	**2.08**	6.00	19.53	8.63	3.59	**4.14**
	Green	5.97	1.71	**1.36**	2.10	6.16	11.68	2.67	**1.60**
	Blue	2.96	1.43	**0.93**	1.19	22.98	18.45	0.96	**1.06**
ytma23_022103	Red	2.80	5.44	1.73	5.35	12.69	4.95	**1.37**	**4.00**
	Green	2.16	1.52	1.34	1.80	5.42	12.44	**1.07**	**1.51**
	Blue	0.67	1.76	0.73	1.38	21.09	18.08	**0.58**	**0.77**
ytma49_042003	Red	3.20	3.19	3.34	2.94	18.85	4.46	**2.66**	**2.94**
	Green	1.81	**0.49**	2.17	0.57	5.65	11.14	1.72	0.61
	Blue	0.91	1.50	0.60	1.26	23.14	18.93	**0.55**	**0.83**
ytma49_042203	Red	9.36	4.16	**2.38**	3.20	7.32	7.01	2.80	**2.77**
	Green	4.98	9.62	**1.85**	0.79	4.81	3.14	2.24	**0.58**
	Blue	16.37	13.66	**0.47**	2.07	0.99	9.39	0.54	**2.06**
ytma49_042403	Red	3.48	3.53	**1.23**	3.74	17.89	4.91	2.22	**2.30**
	Green	2.37	**0.92**	**0.72**	1.06	5.92	11.39	1.16	0.93
	Blue	0.76	1.53	0.62	0.94	22.68	19.31	**0.49**	**0.87**
ytma49_072303	Red	2.11	5.00	1.15	3.71	1.04	1.90	**1.04**	**1.73**
	Green	1.66	1.39	0.95	1.38	0.84	**0.68**	**0.73**	0.76
	Blue	0.34	1.55	0.31	1.04	0.27	0.63	**0.20**	**0.40**
ytma49_111003	Red	**1.99**	**3.53**	2.68	5.62	19.34	6.61	2.74	3.60
	Green	**1.56**	**1.36**	2.44	2.91	2.24	9.68	2.61	1.58
	Blue	1.07	1.82	0.83	2.00	23.70	18.82	**0.77**	**1.20**
ytma49_111303	Red	2.22	4.96	2.01	6.55	19.98	6.91	**1.38**	**3.35**
	Green	2.15	1.95	1.91	2.64	4.96	12.02	**1.31**	**1.69**
	Blue	0.65	1.53	0.79	2.02	25.99	19.91	**0.59**	**1.17**
ytma55_030603	Red	7.08	6.75	**2.08**	4.58	2.36	8.33	2.99	**3.60**
	Green	6.01	3.19	**1.75**	1.78	2.04	1.62	2.78	**1.46**
	Blue	1.15	9.99	1.01	2.06	0.86	4.09	**0.85**	**1.63**

stain and E-stain, respectively. However, the PF, EPF, HTN, SPCN, EM and RFCC$_{cosine}$, respectively, provide lowest values of σ in only 0, 2, 6, 1, 5 and 4 cases for H-stain and 4, 8, 2, 1, 1 and 1 cases for E-stain. All the results reported in Table 4.5 confirm that the proposed algorithm performs better than the existing blind stain estimation methods.

Table 4.4: Comparative Performance Analysis of Different Hue Histograms on CMU Data Set: A good stain separation method should produce lower value of symmetric KL divergence and higher value of SNR. The lowest value of KL divergence and highest value of SNR are marked in bold font.

Image Number	Stain Channel	Standard		SW		α-controlled		RFCC$_{vM}$	
		KL Div	SNR	KL Div	SNR	KL Div	SNR	KL Div	SNR
Image_1	H	2.1251	**22.4391**	1.0561	21.4924	1.3167	21.1020	**1.0245**	22.1049
	E	1.0881	23.5018	0.9988	24.3972	0.9856	**24.5105**	**0.7751**	24.5022
Image_2	H	3.0442	18.0152	3.6845	16.6264	1.5864	20.3162	**1.1631**	**20.3226**
	E	0.8626	21.4878	1.0089	19.8965	0.9923	22.2139	**0.8244**	**22.4269**
Image_3	H	1.9818	19.9786	1.7686	22.0074	1.9008	22.4536	**1.3476**	**22.7224**
	E	**0.6986**	23.4881	0.9335	23.4804	1.0151	23.2550	0.7053	**23.5835**
Average	H	2.3837	20.1443	2.1698	20.0421	1.6013	21.2906	**1.1784**	**21.7166**
	E	0.8831	22.8259	0.9804	22.5914	0.9977	23.3265	**0.7683**	**23.5042**
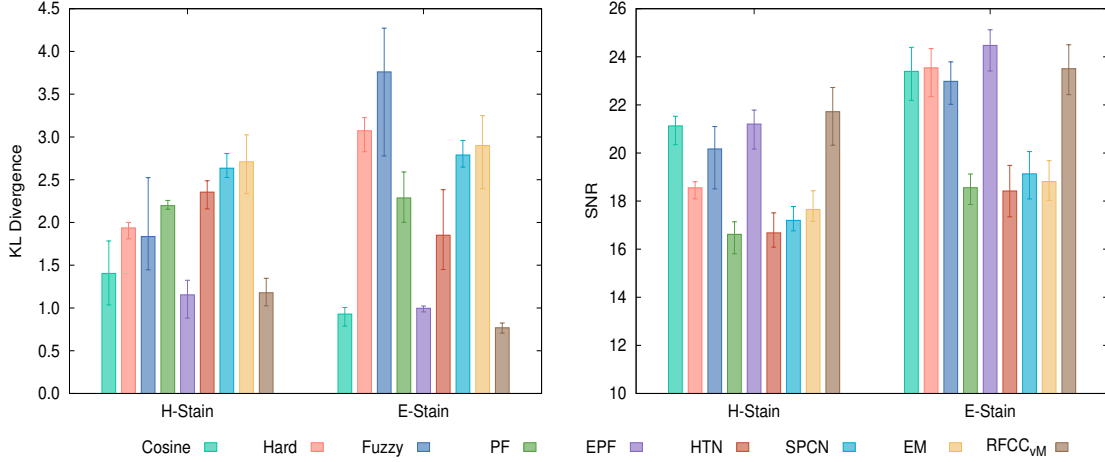
Figure 4.4: Performance analysis of different clustering algorithms, dissimilarity measures, and stain separation algorithms on CMU data set using mean value of each index with error bar: A good stain separation method should produce lower value of symmetric KL divergence and higher value of SNR.

4.3.2 Evaluation of Stain Separation

To establish the effectiveness of the proposed method in stain separation, extensive experiment is performed on CMU data set as it has stain decomposition ground-truth corresponding to each image. The associated results are reported in Fig. 4.4 with respect to symmetric KL divergence and SNR. Fig. 4.4 shows the mean value of each quantitative index along with error bar for each of the algorithms compared. The mean, maximum, and minimum values of each index are calculated for individual stain, considering three H&E stained images of CMU data set. From the results reported in Fig. 4.4, it is found that the proposed dissimilarity measure attains lower mean values of symmetric KL divergence and higher mean values of SNR for both the stains, with respect to cosine distance. Also, the

Table 4.5: Comparative Performance Analysis of Different Algorithms on UCSB Data Set: σ_X (in 10^{-2} scale) represents the standard deviation of the estimated stain vectors of X-stain, computed over all images of the same biopsy set, and $X \in \{H, E\}$. Here, 'Ch.' denotes different channels and $RFCC_c$ represents the method $RFCC_{cosine}$ [104].

Biopsy Sets	Ch.	PF		HTN		EPF		SPCN		EM		RFCC$_c$		RFCC$_{vM}$	
		σ_H	σ_E	σ_H	σ_E	σ_H	σ_E	σ_H	σ_E	σ_H	σ_E	σ_H	σ_E	σ_H	σ_E
ytma10_010704	R	3.04	2.72	2.67	3.25	2.49	**2.10**	2.42	3.93	2.45	4.10	**2.39**	3.65	2.84	2.35
	G	2.23	2.01	1.68	0.97	2.17	1.03	1.66	0.73	1.61	**0.69**	1.39	1.03	**1.27**	0.89
	B	2.67	5.79	1.14	2.43	**0.78**	2.90	1.01	2.55	0.92	2.51	1.59	2.55	1.64	**1.86**
ytma12_010804	R	1.99	**3.79**	**1.37**	5.79	2.02	8.31	2.20	6.26	2.08	7.73	1.89	5.73	3.59	4.14
	G	2.16	2.82	**1.07**	1.82	2.20	**1.13**	1.80	2.14	1.66	2.54	1.17	2.03	2.67	1.60
	B	1.72	6.81	**0.55**	1.30	0.72	1.58	0.79	1.36	0.83	1.56	0.94	1.19	0.96	**1.06**
ytma23_022103	R	1.77	**3.43**	2.39	5.36	2.33	6.73	**1.07**	5.75	1.18	7.27	1.67	5.31	1.37	4.00
	G	1.95	**0.46**	2.25	1.59	2.62	0.64	1.10	1.81	1.16	2.21	1.24	1.80	**1.07**	1.51
	B	1.31	3.34	0.70	1.57	0.92	1.50	0.86	1.71	0.89	1.86	0.64	1.38	**0.58**	**0.77**
ytma49_042003	R	3.11	2.11	3.17	1.86	3.67	**0.55**	3.14	2.98	4.56	4.53	3.28	2.36	**2.66**	2.94
	G	2.98	2.02	2.35	**0.45**	3.50	0.45	2.44	0.74	3.40	0.96	1.99	0.54	**1.72**	0.61
	B	1.09	3.67	0.55	0.98	1.42	1.39	0.79	1.18	1.21	1.42	0.69	0.94	**0.55**	**0.83**
ytma49_042203	R	2.29	**2.18**	4.04	3.13	2.13	5.97	1.99	3.36	**1.93**	3.88	2.24	3.07	2.80	2.77
	G	2.95	1.40	3.81	0.92	2.43	1.21	2.02	0.84	2.06	1.18	**1.49**	0.87	2.24	**0.58**
	B	1.88	3.72	1.15	2.12	0.97	2.99	0.60	2.03	0.69	2.02	0.64	**1.89**	**0.54**	2.06
ytma49_042403	R	3.08	3.13	3.42	3.67	3.03	**0.49**	2.99	4.89	3.39	6.66	**1.72**	3.80	2.22	2.30
	G	3.16	3.18	2.55	0.96	3.09	**0.35**	2.38	1.38	2.72	1.67	**0.94**	1.10	1.16	0.93
	B	1.04	5.38	0.63	**0.47**	0.79	1.09	0.93	0.64	0.92	0.80	0.76	0.72	**0.49**	0.87
ytma49_072303	R	2.09	2.63	2.49	2.82	2.73	2.60	1.99	**1.57**	2.52	1.99	1.41	2.86	**1.04**	1.73
	G	2.48	1.22	2.55	0.95	3.39	**0.14**	2.15	0.66	2.67	0.95	1.14	1.09	**0.73**	0.76
	B	0.57	1.43	0.38	0.80	0.64	0.79	0.46	0.79	0.36	0.98	0.30	0.67	**0.20**	**0.40**
ytma49_111003	R	1.90	13.17	1.64	6.94	5.66	14.97	4.00	9.65	**1.62**	8.12	2.79	5.34	2.74	**3.60**
	G	2.60	11.29	1.61	3.49	7.44	14.24	1.69	5.79	**1.45**	4.63	2.39	2.69	2.61	**1.58**
	B	3.62	10.73	1.23	2.79	4.37	12.09	9.77	4.17	1.29	2.62	1.08	1.93	**0.77**	**1.20**
ytma49_111303	R	2.15	6.12	**1.20**	6.71	3.39	10.67	1.44	6.85	1.63	8.87	2.29	6.09	1.38	**3.35**
	G	2.52	1.75	**1.30**	2.46	4.45	1.78	1.56	2.85	1.75	3.27	2.06	2.54	1.31	**1.69**
	B	1.50	4.30	**0.24**	2.34	1.37	2.49	0.34	2.31	0.55	2.70	0.80	1.80	0.59	**1.17**
ytma55_030603	R	2.18	7.85	1.88	16.56	2.41	5.76	1.35	5.22	**1.25**	7.60	1.91	3.84	2.99	**3.6**
	G	3.39	3.90	1.81	2.09	3.34	**1.34**	1.48	2.36	**1.37**	3.19	1.52	1.62	2.78	1.46
	B	2.73	8.20	0.67	8.32	**0.53**	**1.57**	0.70	1.62	0.67	2.28	0.77	1.67	0.85	1.63

proposed rough-fuzzy clustering algorithm provides higher mean SNR and lower values of KL divergence in all the cases, with respect to both hard and fuzzy clustering. Moreover, the difference among mean, minimum, and maximum values of each index for the proposed dissimilarity measure and rough-fuzzy clustering is significantly smaller as compared to that of other approaches.

In order to establish the relevance of weighted hue histogram in stain separation over standard hue, saturation-weighted (SW) hue, and α-controlled hue histograms, extensive
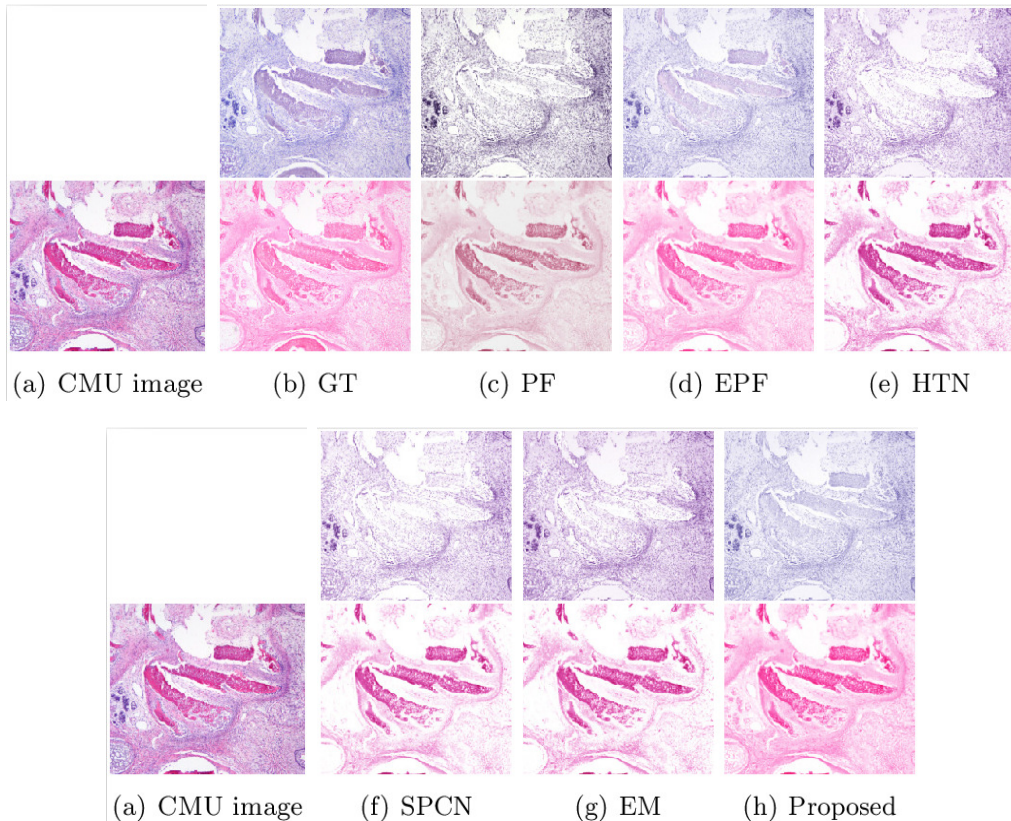
Figure 4.5: CMU image_1, ground-truth stain spectra, and estimated stain spectra obtained using different stain separation methods (top: H-stain; bottom: E-stain). In the figure, 'Proposed' method refers to the proposed $RFCC_{vM}$ method.

experimentation is done on CMU data set, and corresponding results are reported in Table 4.4. Table 4.4 compares the performance of different hue histograms on CMU data set. Results are reported with respect to symmetric KL divergence and SNR. The decomposition ground-truth of both H-stained and E-stained images are available, corresponding to each image of the CMU data set. So, the decomposition result, obtained for each image of CMU data set, is compared with the decomposition ground-truth to compute both KL divergence and SNR. All the results reported in Table 4.4 confirm that the proposed hue histogram attains lowest values of KL divergence and highest values of SNR in 7 and 6 cases, respectively, out of 8 cases each.

The results reported in Fig. 4.4 also confirm that the proposed method provides lower values for KL divergence and higher values of SNR, with respect to most of the existing stain separation algorithms, while EPF performs slightly better than the proposed approach, with respect to KL divergence for H-stain and SNR for E-stain. Fig. 4.5 depicts an example of qualitative comparison of stain separation by different methods. From Fig. 4.5, it is evident that only the proposed method and EPF can extract the intrinsic structures of the biological components, highlighted by H-stain and E-stain. But, the main advantage of the proposed approach over EPF is that it is also applicable in the situation where more than two stains exist.
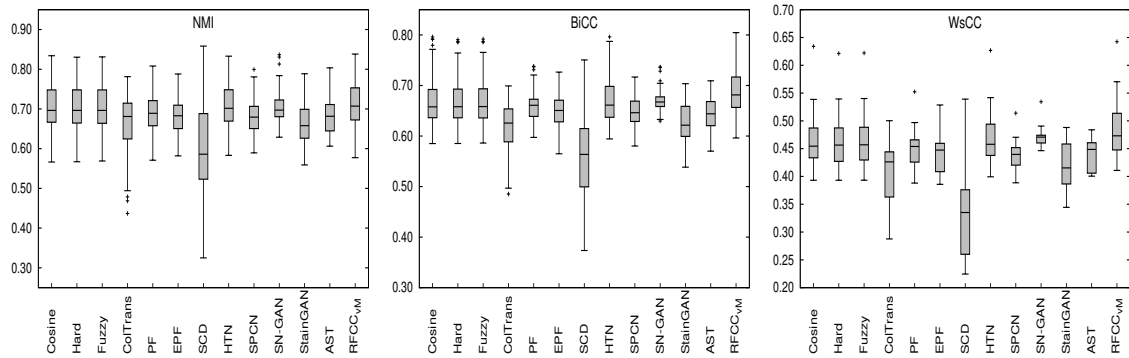
Figure 4.6: Performance of different clustering algorithms, dissimilarity measures, and color normalization methods on UCSB data set using box and whisker plots for NMI, BiCC, and WsCC: A good color normalization method should produce higher values of these three indices.
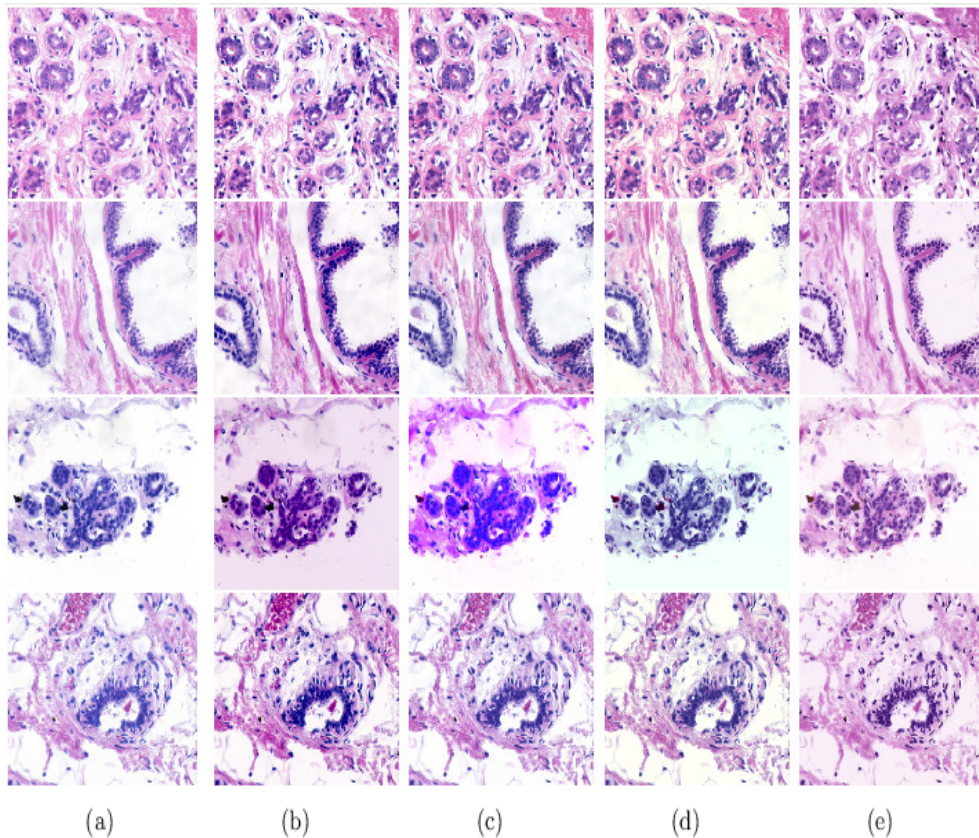
Figure 4.7: (a) Original images of UCSB data set; and color normalized images obtained using different color normalization methods: (b) ColTrans, (c) PF, (d) EPF and (e) SCD.
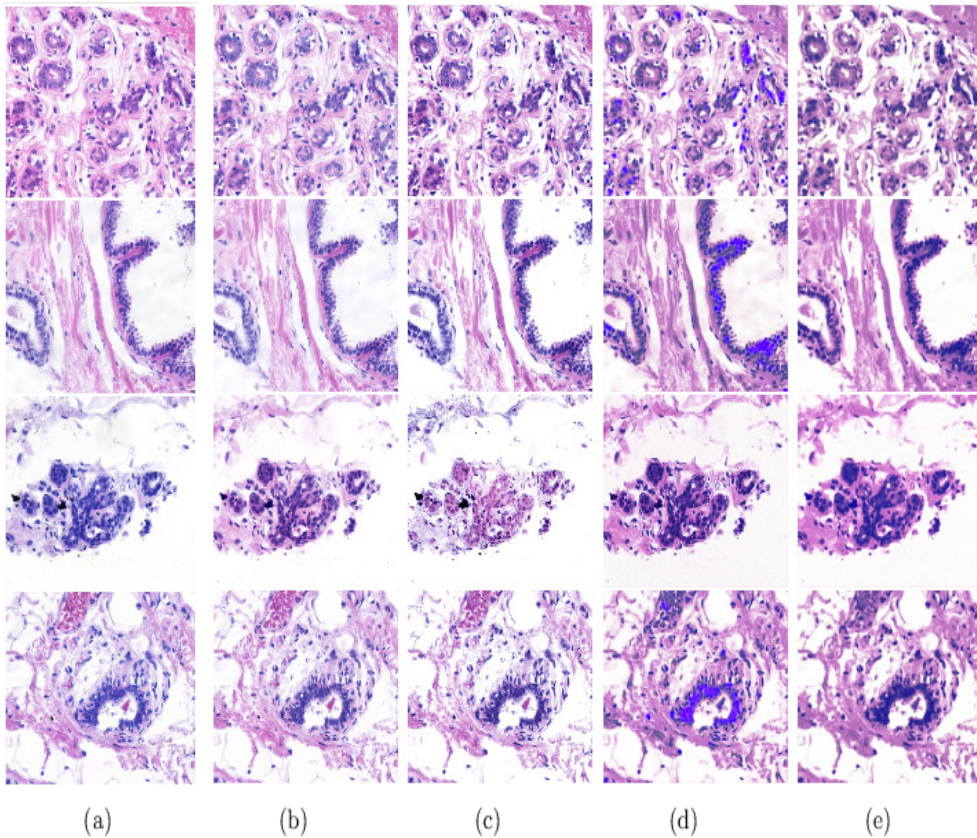
Figure 4.8: (a) Original images of UCSB data set; and color normalized images obtained using different color normalization methods: (b) HTN, (c) SPCN, (d) SN-GAN and (e) StainGAN.

4.3.3 Evaluation of Color Constancy

The within-image and within-biopsy set color consistency of the proposed method as well as other color normalization methods are evaluated on UCSB data set using existing NMI and two proposed indices, namely, BiCC and WsCC. The comparative performance of different methods is analyzed using box and whisker plots, and p-values computed through both paired-t (one-tailed) and Wilcoxon signed-rank (one-tailed) tests. Fig. 4.6 presents the box and whisker plots of three quantitative indices, where the top and bottom boundaries of each box represent upper and lower quartiles, respectively, central line represents the median, whiskers are extended to three standard deviations from mean, and the outliers are represented by '+'. The corresponding statistical significance analysis, with respect to both Wilcoxon signed-rank and paired-t tests, is presented in Table 4.6.

From the results reported in Fig. 4.6 and Table 4.6, it is evident that the proposed circular dissimilarity measure provides higher mean and median values of NMI, BiCC, and WsCC as compared to that of cosine distance. Also, the performance of the proposed dissimilarity measure is significantly better than the cosine distance, with respect to both paired-t and Wilcoxon signed-rank tests considering 95% confidence level. In order to establish the significance of rough-fuzzy clustering over its hard and fuzzy counter parts,
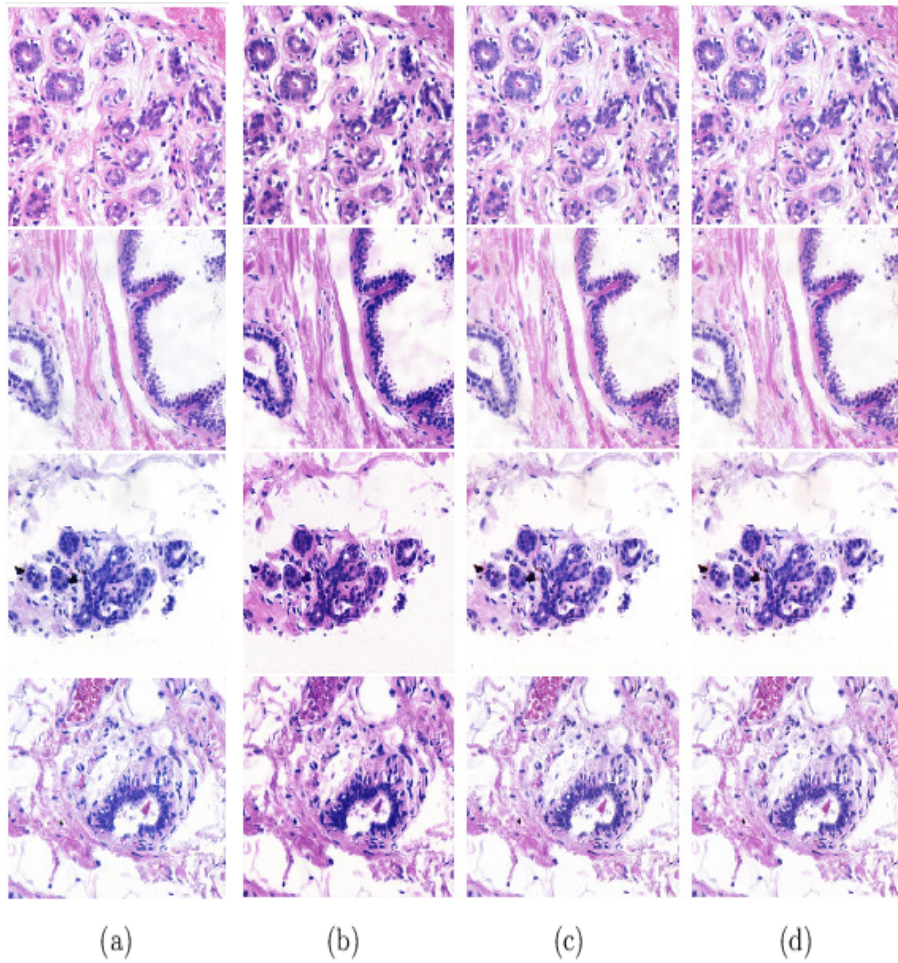
Figure 4.9: (a) Original images of UCSB data set; and color normalized images obtained using different color normalization methods: (b) AST, (c)RFCC$_{cosine}$ and (d) RFCC$_{vM}$.

the comparative results are also reported in Fig. 4.6 and Table 4.6. From the results, it can be seen that the proposed rough-fuzzy circular clustering algorithm provides significantly higher NMI, BiCC, and WsCC values, compared to both hard and fuzzy clustering, on UCSB data set. The significantly better performance of the proposed algorithm is achieved due to the fact that the theory of rough sets can efficiently deal with uncertainty, vagueness and incompleteness in stain class definition.

In order to establish the relevance of weighted hue histogram in color normalization of histological images over standard hue, saturation-weighted (SW) hue, and α-controlled hue histograms, extensive experimentation is done on UCSB data set, and corresponding results are reported in Table 4.6. Table 4.6 represents the comparative performance analysis of different hue histograms on UCSB data set with respect to NMI, BiCC, and WsCC. The comparative performance analysis of different hue histograms is also studied using tables of means, medians, and p-values computed through both paired-t (one-tailed) and Wilcoxon signed-rank (one-tailed) tests. From all the results reported in Table 4.6, it is seen that the proposed weighted hue histogram provides highest mean and median values for NMI, BiCC,

Table 4.6: Statistical Significance Analysis of Different Clustering Algorithms, Dissimilarity Measures, and Color Normalization Methods, with respect to Proposed Method, on UCSB Data: Both paired-t and Wilcoxon signed-rank tests are used. The p-value, marked in italics, indicates the proposed method is better than the existing one, but not significantly.

Different Methods	NMI		BiCC		WsCC	
	Wilcoxon	Paired-t	Wilcoxon	Paired-t	Wilcoxon	Paired-t
Cosine Distance	2.71E-11	1.35E-13	1.75E-11	2.71E-17	2.53E-03	5.17E-06
Hard Clustering	3.27E-11	8.73E-13	1.75E-11	2.28E-19	2.53E-03	5.10E-07
Fuzzy Clustering	1.75E-11	9.00E-12	1.75E-11	2.13E-18	2.53E-03	2.30E-06
ColTrans	3.13E-06	2.45E-06	4.94E-11	1.09E-13	3.46E-03	3.81E-03
PF	3.43E-04	1.52E-04	8.73E-07	3.49E-08	*1.01E-01*	2.67E-02
EPF	1.08E-05	9.56E-07	7.76E-09	1.22E-10	1.42E-02	8.70E-03
SCD	1.43E-10	1.71E-14	1.75E-11	3.22E-22	2.53E-03	2.37E-05
HTN	4.46E-07	4.87E-06	2.16E-11	2.24E-14	2.53E-03	8.75E-05
SPCN	2.91E-06	4.66E-07	1.81E-08	1.76E-10	1.09E-02	7.95E-03
SN-GAN	*3.39E-01*	*2.30E-01*	1.02E-03	2.09E-04	*1.93E-01*	*1.51E-01*
StainGAN	1.80E-06	1.55E-07	1.01E-10	5.93E-14	3.46E-03	2.22E-03
AST	1.53E-03	6.18E-04	3.92E-09	7.46E-11	1.42E-02	9.66E-03

and WsCC as well as performs significantly better than other hue histograms, considering 0.05 as the level of significance. The better performance of the weighted hue histogram is achieved due to the fact that it can reduce the effect of achromatic pixels on saturation-weighted hue histogram as well as it incorporates the local neighborhood information of a pixel in standard hue image.

Finally, the results reported in Fig. 4.6 and Table 4.6 ensure that the proposed method provides highest mean and median values, with respect to NMI, BiCC, and WsCC. Moreover, both Wilcoxon and paired-t tests show that the proposed method performs significantly better than state-of-the-art color normalization algorithms except SN-GAN [41]. The proposed method provides better performance than SN-GAN [41], but not significantly, with respect to both Wilcoxon and paired-t tests in case of NMI and WsCC. The qualitative performance analyses of different methods on UCSB data set are presented in Fig. 4.7, Fig. 4.8 and Fig. 4.9. From the results reported in Table 4.6 and Fig. 4.6-4.9, it can be observed that the proposed method outperforms other existing methods as per the color consistency after normalization is concerned.

4.4 Conclusion

The color normalization in histological images is a fundamental task as the presence of color inconsistency among the images may degrade the performance of automated histological image analysis. In this regard, the main contributions of this chapter are as follows:

- development of a new rough-fuzzy circular clustering algorithm for color normalization, which judiciously integrates the merits of rough sets and fuzzy set;

- introducing a new dissimilarity measure to deal with the circular nature of hue values;

- introducing a weighted hue histogram, integrating the information of both saturation and local neighborhood;

- defining some new quantitative indices to evaluate the color constancy among H&E stained histological images after color normalization; and

- demonstrating the efficacy of the proposed method, along with a comparison with existing algorithms, on publicly available standard H&E stained histological images.

The proposed method addresses the physical constraints such as the overlapping nature of histochemical stains as well as the non-negativity of stain color appearance and stain density matrices. It is evident from both quantitative and qualitative results that the proposed method outperforms other existing color normalization approaches, in terms of satisfying the physical constraints and maintaining the within-image and between-image color constancy within a single biopsy set after color normalization. Moreover, the proposed algorithm does not require any prior information other than the number of stains involved in staining routine of histological images.

The rough-fuzzy circular clustering based stain color normalization methods, proposed in Chapter 3 and this chapter, do not explore the correlation between color appearance and stain bound information, corresponding to each histological image. Hence, it can not verified that the loss or modification in one information does not affect the other one, which may eventually lead to loss of histological information after color normalization. In order to normalize the color appearance without hampering the histological information of the image, the next chapter introduces a new deep generative model to reduce the color variation present among the histological images.

Chapter 5

Truncated Normal Mixture Prior Based Deep Latent Model for Color Normalization of Histological Images

5.1 Introduction

Histological image analysis deals with the microscopic examination of tissue slides for the diagnosis of diseases under consideration. The most important characteristic of histological data is the massive amount of information, which makes computer aided analysis more accurate than cytological, radiological and other imaging modalities. Tissue slides are stained with different histochemical reagents, which adhere to distinct cellular protein structures, and having contrasting appearances, highlight different cellular components. The relevance of this color information in quantitative study of histology slides has been established by some methods [75].

One of the foremost and significant problems of histological image analysis is the un-acceptable inter-sample, intra-sample and inter-biopsy disagreement in the appearance of stained tissue color. As a result, the numerical features extracted from non-normalized images may lead to faulty image interpretation by computer aided diagnosis systems [15]. The stain separation based approaches try to investigate the actual causes for color dis-agreement. In this respect, some state-of-the-art color normalization methods, such as, [8], [15], [84], [86], [132], [133] and [22], have already been discussed in Chapter 3 and Chapter 4.

The stain bound information, which depicts the adhesiveness of multiple contrasting stains to different cellular protein structures, must be kept intact after color normalization. If the color appearance information is captured accurately, then the information extracted from template image can be utilized to perform the color normalization. Due to the co-occurrence of constituent stains, one of the major challenges of color normalization is color mixing due to overlapping nature of chemical stains [134]. An expectation-maximization (EM) approach in circular hue domain has been proposed in [134] to estimate the stain concentration matrix. Recently, a rough-fuzzy circular clustering (RFCC_{vM}) based color

normalization method has been proposed in [105]. Although these two approaches [105,134] have tried to address the overlapping characteristics of different stain classes, both of them use approximated hue histograms, which fails to capture spatial relationship. Moreover, they rely on NMF, which suffers from unstable convergence problem, to satisfy nonnegativity constraint corresponding to stain color appearance matrix and associated stain density map.

In recent years, deep learning has emerged as an effective tool in the field of medical image analysis. In [6], a task-specific discriminative image analysis model has been proposed to learn color transformation and to extract staining information. However, its performance is hugely dependent on the auxiliary tasks that require extra labeling efforts. A generative adversarial network (GAN) based end-to-end learning model has been proposed in [41], to capture the color appearances of different stains involved. Being inspired by the unpaired image-to-image translation model CycleGAN [64], StainGAN method [90] has included cycle-consistency loss in the objective function. Recently, a capsule network based stain standardization model has been proposed in [145] for the selection of most appropriate parameters for stain separation. A GPU enabled color normalization method for digital pathology has been proposed in [32]. Recently, a stain normalization method has been proposed in [27] based on texture enhanced GANs. In [18], a fast self-supervised fully convolutional network has been proposed for color normalization of histology images. Using self-supervised semantic guidance, a structure-preserving stain normalization method has been proposed in [39]. The major limitations of the above methods are the inability to deal with the uncertainty due to overlapping nature of histochemical stains, and inability to extract stain color appearance information as well as stain bound information corresponding to each histology image.

In this regard, this chapter introduces a new deep generative model, termed as TredMiL (Truncated normal Mixture prior based Latent model), for color normalization of histology images. The model assumes that the latent color appearance code and stain density code are independent of each other. A generative module and a reconstructive module are designed accordingly to capture disentangled color appearance and stain density information. The disentangled representation enhances the generalizability and adaptablity of the model in the situation where the loss or modification in one information does not affect the other information. To deal with the overlapping nature of histochemical reagents, the proposed model assumes that the latent color appearance code, extracted through the color appearance encoder, is sampled from a mixture model. As the outer tails of a mixture model do not contribute adequately in handling overlapping information, rather are prone to outliers, a mixture of truncated normal distributions is used to deal with the overlapping nature of stains. The performance of the proposed model, along with a comparison with state-of-the-art approaches, is demonstrated on several publicly available data sets containing H&E stained histological images. Some of the results presented in this chapter are reported in [118].

The the rest of this chapter is organized as follows: Section 5.2 introduces a novel deep generative model, along with the theoretical details regarding generative module as well as reconstructive module, to address the stain color normalization problem of histological images. The efficacy of the proposed deep generative model is analyzed with respect to several existing stain separation and color normalization approaches on publicly available histological image data sets in Section 5.3. Finally, the chapter ends with the concluding

remarks provided in Section 5.4.

5.2 TredMiL: Proposed Method

This section presents a deep generative model for color normalization of histology images. The proposed model is developed based on least square GAN (LSGAN) model [136].

5.2.1 Problem Statement

Given a histological image set S, consisting of N biopsy sets $\{x^i\}$, $i = 1, 2, \cdots, N$, the goal is to build a model that takes a non-normalized histological image as input and outputs the corresponding color normalized image. As the biopsy set images are collected from different subjects and laboratories, it is clearly evident that the biopsy images of two sets $\{x^i\}$ and $\{x^j\}$, $i \neq j$, differ in stain color appearance due to several factors associated with color disagreement. Moreover, any arbitrary pair of images x_p^i and x_q^i, $p \neq q$, of the same biopsy set $\{x^i\}$, exhibits color variation due to variation in specimen width. Hence, the aim of the current study is to develop a deep framework that, with proper training, will be able to generate color normalized histological image corresponding to each non-normalized histological image slide.
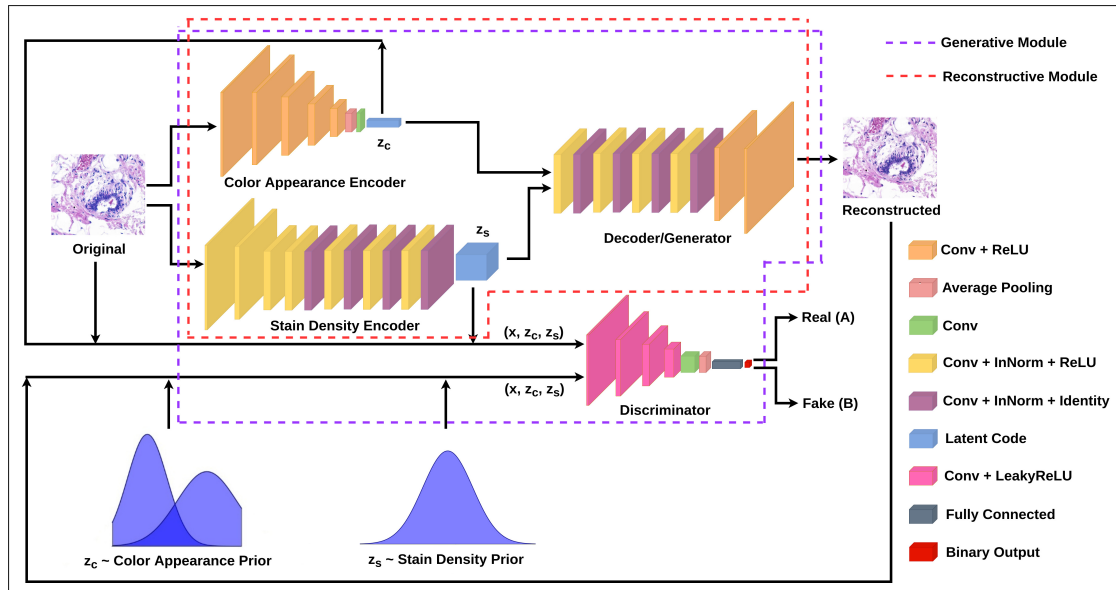
Figure 5.1: Block diagram of the proposed deep generative model TredMiL for color normalization of histological images. Here, each network component is marked with distinct color representation. In the diagram, 'Conv' means convolutional layer, 'ReLU' denotes rectified linear unit, 'InNorm' represents instance normalization, 'Identity' means identity function, i.e., $f(x) = x$, 'LeakyReLU' denotes leaky rectified linear unit. Generative module and reconstructive module are marked with violet dashed line and red dashed line, respectively. The network architecture follows image-to-image translation model MUNIT [131]. The complete model is trained end-to-end using backpropagation technique.

5.2.2 Model Structure

Different components of the proposed TredMiL model is shown in Fig. 5.1. The proposed model consists of four deep networks: a color appearance encoder $\mathcal{E}_c$, which extracts the color appearance information, a stain density encoder $\mathcal{E}_s$, which captures the information regarding amount of stains bound to biological components, a decoder/generator $\mathcal{G}$ and a discriminator $\mathcal{D}$. Each training set image x is fed simultaneously into both the color appearance encoder $\mathcal{E}_c(x; \theta_{\mathcal{E}_c})$ and the stain density encoder $\mathcal{E}_s(x; \theta_{\mathcal{E}_s})$, which eventually output the latent color appearance code z_c and latent stain density representation z_s, respectively. Both the latent representations, z_c and z_s, are then fed into the decoder/generator $\mathcal{G}(z_c, z_s; \theta_{\mathcal{G}})$ as inputs, which generates the reconstructed image $\hat{x}$. The discriminator $\mathcal{D}(x, z_c, z_s; \theta_{\mathcal{D}})$ takes inputs in the form of triplets (x, z_c, z_s) and discriminates real encoding, where x comes from real data distribution, z_c and z_s correspond to latent codes obtained through $\mathcal{E}_c$ and $\mathcal{E}_s$, respectively, from generated/fake encoding, where x denotes the reconstructed sample $\hat{x}$ and z_c, z_s come from corresponding prior distributions. Here, $\theta_{\mathcal{E}_c}$, $\theta_{\mathcal{E}_s}$, $\theta_{\mathcal{G}}$ and $\theta_{\mathcal{D}}$ represent the network parameters associated with the deep networks $\mathcal{E}_c$, $\mathcal{E}_s$, $\mathcal{G}$ and $\mathcal{D}$, respectively. The networks $\mathcal{E}_c$, $\mathcal{E}_s$, $\mathcal{G}$ and $\mathcal{D}$ should be any differentiable functions so that the error values can be back-propagated during training. The underlying networks are chosen to be convolutional neural networks considering their tremendous success in different image analysis tasks.

5.2.3 Objective Function

The objective of the model is to extract disentangled latent color appearance information and stain bound information, and to learn the generated/fake encoding as close as possible to the real encoding by minimizing the following objective function:

$$J_{\text{total}} = \lambda_{adv} \times J_{\text{adv}} + \lambda_{rec} \times J_{\text{rec}}, \qquad (5.1)$$

where the minimization objective terms J_{adv} and J_{rec} are attributed by the generative module and the reconstructive module, respectively, which follow next. Here, λ_{adv} and λ_{rec} denote the relative importance of the adversarial objective J_{adv} and the reconstruction objective J_{rec}, respectively. In the current study, J_{total} is considered to be the convex combination of J_{adv} and J_{rec}, that is, $\lambda_{adv} + \lambda_{rec} = 1$.

5.2.3.1 Generative Module

The generative module is comprised of all four deep networks: $\mathcal{E}_c$, $\mathcal{E}_s$, $\mathcal{G}$ and $\mathcal{D}$, as marked in Fig. 5.1 with violet dashed line. The generative module incorporates the assumption that the latent color appearance information and latent stain bound information are independent of each other. The objective functions, corresponding to the generative module, are given as follows:

$$J_{\text{G}}(\mathcal{D}) = \min_{\mathcal{D}} J_1(\mathcal{E}_c, \mathcal{E}_s, \mathcal{G}, \mathcal{D}), \qquad (5.2)$$

where $J_1(\mathcal{E}_c, \mathcal{E}_s, \mathcal{G}, \mathcal{D}) = \underbrace{E_{x \sim P_X(x)} E_{z_c \sim P_{\mathcal{E}_c}(z_c|x)} E_{z_s \sim P_{\mathcal{E}_s}(z_s|x)} (A - \mathcal{D}[x, z_c, z_s])^2}_{R} +$

$$\underbrace{E_{z_c \sim P_{\mathcal{Z}_c}(z_c)} E_{z_s \sim P_{\mathcal{Z}_s}(z_s)} E_{x \sim P_{\mathcal{G}}(x|z_c,z_s)} (B - \mathcal{D}[x, z_c, z_s])^2}_{G},$$

A and B denote the labels, assigned by discriminator $\mathcal{D}$, to designate real and generated/fake encoding, respectively, and

$$J_{\mathrm{G}}(\mathcal{G}) = \min_{\mathcal{G}} J_2(\mathcal{E}_c, \mathcal{E}_s, \mathcal{G}, \mathcal{D}),$$

where $J_2(\mathcal{E}_c, \mathcal{E}_s, \mathcal{G}, \mathcal{D}) = E_{z_c \sim P_{\mathcal{Z}_c}(z_c)} E_{z_s \sim P_{\mathcal{Z}_s}(z_s)} E_{x \sim P_{\mathcal{G}}(x|z_c,z_s)} (C - \mathcal{D}[x, z_c, z_s])^2$ (5.3)

C represents the label, assigned by discriminator $\mathcal{D}$ to designate generated/fake encoding, as desired by generator $\mathcal{G}$.

In part R of (5.2), $P_X(x)$ represents the real data distribution, $P_{\mathcal{E}_c}(z_c \mid x)$ and $P_{\mathcal{E}_s}(z_s \mid x)$ denote the conditional distributions corresponding to the encoders $\mathcal{E}_c$ and $\mathcal{E}_s$, respectively, given the data sample x. In part G of (5.2), $P_{\mathcal{Z}_c}(z_c)$ and $P_{\mathcal{Z}_s}(z_s)$ denote the prior distributions corresponding to latent color appearance code z_c and stain density code z_s, respectively, and $P_{\mathcal{G}}(x \mid z_c, z_s)$ represents the conditional distribution corresponding to generated sample $\hat{x}$, given the latent codes z_c and z_s. Now, (5.2) can be simplified as follows:

$$J_1(\mathcal{E}_c, \mathcal{E}_s, \mathcal{G}, \mathcal{D}) = \int_x P_X(x) \int_{z_c} P_{\mathcal{E}_c}(z_c \mid x) \int_{z_s} P_{\mathcal{E}_s}(z_s \mid x) \times (A - \mathcal{D}[x, z_c, z_s])^2 dz_s dz_c dx$$

$$+ \int_{z_c} P_{\mathcal{Z}_c}(z_c) \int_{z_s} P_{\mathcal{Z}_s}(z_s) \int_x P_{\mathcal{G}}(x \mid z_c, z_s)(B - \mathcal{D}[x, z_c, z_s])^2 dx dz_s dz_c$$

$$= \int_{\{x, z_c, z_s\}} [P_X(x) P_{\mathcal{E}_c}(z_c \mid x) P_{\mathcal{E}_s}(z_s \mid x)(A - \mathcal{D}[x, z_c, z_s])^2$$

$$+ P_{\mathcal{Z}_c}(z_c) P_{\mathcal{Z}_s}(z_s) P_{\mathcal{G}}(x \mid z_c, z_s)(B - \mathcal{D}[x, z_c, z_s])^2] dx dz_c dz_s$$

$$= \int_{\{x, z_c, z_s\}} [P_X(x) P_{\mathcal{E}_c \mathcal{E}_s}(z_c, z_s \mid x)(A - \mathcal{D}[x, z_c, z_s])^2 +$$

$$P_{\mathcal{Z}_c, \mathcal{Z}_s}(z_c, z_s) P_{\mathcal{G}}(x \mid z_c, z_s)(B - \mathcal{D}[x, z_c, z_s])^2] dx dz_c dz_s$$

as latent codes z_c and z_s are assumed to be independent. So,

$$J_1(\cdot) = \int_{\{x, z_c, z_s\}} [P_{\mathcal{E}_c \mathcal{E}_s X}(x, z_c, z_s)(A - \mathcal{D}[x, z_c, z_s])^2 +$$

$$P_{\mathcal{G} \mathcal{Z}_c \mathcal{Z}_s}(x, z_c, z_s)(B - \mathcal{D}[x, z_c, z_s])^2] dx dz_c dz_s. \tag{5.4}$$

At first, the optimal discriminator $\mathcal{D}$ is to be computed for any given encoder-decoder combination $(\mathcal{E}_c, \mathcal{E}_s, \mathcal{G})$.

Proposition 5.1. *For a fixed combination of color appearance encoder $\mathcal{E}_c$, stain density encoder $\mathcal{E}_s$ and decoder/ generator $\mathcal{G}$, the optimal discriminator $\mathcal{D}$ can be computed as:*

$$\mathcal{D}^*[x, z_c, z_s] = \frac{A.P_{\mathcal{E}_c\mathcal{E}_s X}(x, z_c, z_s) + B.P_{\mathcal{G}Z_c Z_s}(x, z_c, z_s)}{P_{\mathcal{E}_c\mathcal{E}_s X}(x, z_c, z_s) + P_{\mathcal{G}Z_c Z_s}(x, z_c, z_s)} \tag{5.5}$$

Proof. For a fixed combination of $\mathcal{E}_c$, $\mathcal{E}_s$ and $\mathcal{G}$, the learning criterion for discriminator $\mathcal{D}$ is to minimize $J_1(\mathcal{E}_c, \mathcal{E}_s, \mathcal{G}, \mathcal{D})$ with respect to $\mathcal{D}$ by computing partial derivative as follows:

$$\frac{\partial J_1(\mathcal{E}_c, \mathcal{E}_s, \mathcal{G}, \mathcal{D})}{\partial \mathcal{D}[x, z_c, z_s]} = 2P_{\mathcal{E}_c\mathcal{E}_s X}(x, z_c, z_s)(\mathcal{D}[x, z_c, z_s]$$
$$-A) + 2P_{\mathcal{G}Z_c Z_s}(x, z_c, z_s)(\mathcal{D}[x, z_c, z_s] - B) \tag{5.6}$$

Let the optimal discriminator be denoted as $\mathcal{D}^*[x, z_c, z_s]$. So, at optimal point corresponding to the discriminator $\mathcal{D}$:

$$\left. \frac{\partial J_1(\mathcal{E}_c, \mathcal{E}_s, \mathcal{G}, \mathcal{D})}{\partial \mathcal{D}[x, z_c, z_s]} \right|_{\mathcal{D}=\mathcal{D}^*} = 0;$$

$$\Rightarrow 2\{P_{\mathcal{E}_c\mathcal{E}_s X}(x, z_c, z_s)(\mathcal{D}^*[x, z_c, z_s] - A)+$$

$$P_{\mathcal{G}Z_c Z_s}(x, z_c, z_s)(\mathcal{D}^*[x, z_c, z_s] - B)\} = 0; \quad [\text{using } (5.6)]$$

$$\Rightarrow \mathcal{D}^*[x, z_c, z_s] = \frac{A.P_{\mathcal{E}_c\mathcal{E}_s X}(x, z_c, z_s) + B.P_{\mathcal{G}Z_c Z_s}(x, z_c, z_s)}{P_{\mathcal{E}_c\mathcal{E}_s X}(x, z_c, z_s) + P_{\mathcal{G}Z_c Z_s}(x, z_c, z_s)} \tag{5.7}$$

Theoretically, for optimal combination of encoder-decoder $(\mathcal{E}_c, \mathcal{E}_s, \mathcal{G})^*$, the discriminator should not be able to discriminate between real encoding and generated/fake encoding as generator $\mathcal{G}$ learns the intrinsic real joint distribution. As a result, at that point, the optimal discriminator $\mathcal{D}^*$ cannot distinguish between real and generated/fake encoding, and designates both of them by the same label C that the generator $\mathcal{G}$ wants $\mathcal{D}$ to believe. So, from (5.4), it can be written that

$$(\mathcal{E}_c, \mathcal{E}_s, \mathcal{G})^* = \arg \min_{\mathcal{E}_c, \mathcal{E}_s, \mathcal{G}} J_2(\mathcal{E}_c, \mathcal{E}_s, \mathcal{G}, \mathcal{D}^*),$$

where $J_2(\mathcal{E}_c, \mathcal{E}_s, \mathcal{G}, \mathcal{D}^*) = E_{x \sim P_X(x)} E_{z_c \sim P_{\mathcal{E}_c}(z_c|x)} E_{z_s \sim P_{\mathcal{E}_s}(z_s|x)} (C - \mathcal{D}^*[x, z_c, z_s])^2$

$$+ E_{z_c \sim P_{Z_c}(z_c)} E_{z_s \sim P_{Z_s}(z_s)} E_{x \sim P_{\mathcal{G}}(x|z_c, z_s)} (C - \mathcal{D}^*[x, z_c, z_s])^2$$

[adding terms independent of $\mathcal{G}$ does not affect optimization]

$$= \int_{\{x, z_c, z_s\}} \{P_{\mathcal{E}_c\mathcal{E}_s X}(x, z_c, z_s)(C - \mathcal{D}^*[x, z_c, z_s])^2 + P_{\mathcal{G}Z_c Z_s}(x, z_c, z_s)(C - \mathcal{D}^*[x, z_c, z_s])^2\} dx dz_c dz_s$$

$$= \int_{\{x,z_c,z_s\}} \{P_{\mathcal{E}_c\mathcal{E}_s\mathcal{X}}(x, z_c, z_s) + P_{\mathcal{G}z_cz_s}(x, z_c, z_s)\}$$

$$\times \left[C - \frac{A.P_{\mathcal{E}_c\mathcal{E}_s\mathcal{X}}(x, z_c, z_s) + B.P_{\mathcal{G}z_cz_s}(x, z_c, z_s)}{P_{\mathcal{E}_c\mathcal{E}_s\mathcal{X}}(x, z_c, z_s) + P_{\mathcal{G}z_cz_s}(x, z_c, z_s)}\right]^2 dxdz_cdz_s. \tag{5.8}$$

Theorem 5.1. *Given that $A = 1$ and $B = 0$, for optimal combination of encoder-decoder $(\mathcal{E}_c, \mathcal{E}_s, \mathcal{G})^*$, the optimal discriminator $\mathcal{D}^*$ attains a probability value of $\frac{1}{2}$, if and only if $P_{\mathcal{E}_c\mathcal{E}_s\mathcal{X}} = P_{\mathcal{G}z_cz_s}$.*

Proof. (**IF**) If $P_{\mathcal{E}_c\mathcal{E}_s\mathcal{X}} = P_{\mathcal{G}z_cz_s}$, then (5.7) can be rewritten as follows:

$$\mathcal{D}^*[x, z_c, z_s] = \frac{A.P_{\mathcal{E}_c\mathcal{E}_s\mathcal{X}}(x, z_c, z_s) + B.P_{\mathcal{G}z_cz_s}(x, z_c, z_s)}{P_{\mathcal{E}_c\mathcal{E}_s\mathcal{X}}(x, z_c, z_s) + P_{\mathcal{G}z_cz_s}(x, z_c, z_s)}$$

$$= \frac{(A + B)P_{\mathcal{E}_c\mathcal{E}_s\mathcal{X}}(x, z_c, z_s)}{2P_{\mathcal{E}_c\mathcal{E}_s\mathcal{X}}(x, z_c, z_s)} = \frac{(A + B)}{2}. \tag{5.9}$$

Substituting the pre-assumed label values $A = 1$ and $B = 0$, (5.9) leads to the following:

$$\mathcal{D}^*[x, z_c, z_s] = \frac{(1 + 0)}{2} = \frac{1}{2}. \tag{5.10}$$

(**ONLY IF**) Now, (5.8) of Section 5.2.3.1 can be simplified as follows:

$$(\mathcal{E}_c, \mathcal{E}_s, \mathcal{G})^*$$

$$= \int_{\{x,z_c,z_s\}} \frac{\left[\begin{array}{c}(C - A)P_{\mathcal{E}_c\mathcal{E}_s\mathcal{X}}(x, z_c, z_s)+ \\ (C - B)P_{\mathcal{G}z_cz_s}(x, z_c, z_s)\end{array}\right]^2}{\left[P_{\mathcal{E}_c\mathcal{E}_s\mathcal{X}}(x, z_c, z_s) + P_{\mathcal{G}z_cz_s}(x, z_c, z_s)\right]} dxdz_cdz_s$$

$$= \int_{\{x,z_c,z_s\}} \frac{\left[\begin{array}{c}(C - A)[P_{\mathcal{E}_c\mathcal{E}_s\mathcal{X}}(x, z_c, z_s)+ \\ P_{\mathcal{G}z_cz_s}(x, z_c, z_s)]+ \\ (A - B)P_{\mathcal{G}z_cz_s}(x, z_c, z_s)\end{array}\right]^2}{\left[P_{\mathcal{E}_c\mathcal{E}_s\mathcal{X}}(x, z_c, z_s) + P_{\mathcal{G}z_cz_s}(x, z_c, z_s)\right]} dxdz_cdz_s \tag{5.11}$$

The real encoding and the generated/fake encoding are considered to have label values $A = 1$ and $B = 0$, respectively. Now, if $\mathcal{D}^*$ attains a probability value $\frac{1}{2}$, that means the optimal discriminator cannot distinguish between real and generated/fake encoding. As a result, label C is eventually assumed to have a value $\frac{1}{2}$. Hence, (5.11) can be rewritten as

$$(\mathcal{E}_c, \mathcal{E}_s, \mathcal{G})^*$$

$$= \frac{1}{4} \int\limits_{\{x,z_c,z_s\}} \frac{\left[\begin{array}{c} 2P_{\mathcal{G}z_c z_s}(x, z_c, z_s) - \\ [P_{\mathcal{E}_c \mathcal{E}_s \mathcal{X}}(x, z_c, z_s) + \\ P_{\mathcal{G}z_c z_s}(x, z_c, z_s)] \end{array} \right]^2}{[P_{\mathcal{E}_c \mathcal{E}_s \mathcal{X}}(x, z_c, z_s) + P_{\mathcal{G}z_c z_s}(x, z_c, z_s)]} dx dz_c dz_s$$

$$= \frac{1}{4} \chi^2_{Pearson} \left[\begin{array}{c} 2P_{\mathcal{G}z_c z_s}(x, z_c, z_s) \, || \, (P_{\mathcal{E}_c \mathcal{E}_s \mathcal{X}}(x, z_c, z_s) \\ + P_{\mathcal{G}z_c z_s}(x, z_c, z_s)) \end{array} \right], \tag{5.12}$$

where $\chi^2_{Pearson}(.)$ represents the Pearson's chi-squared divergence measure. Pearson's chi-squared divergence is a special case of f-divergence measure. The f-divergence between two continuous probability distributions $\mathcal{M}$ and $\mathcal{N}$ is defined as:

$$D_f(\mathcal{M} \, || \, \mathcal{N}) = \int\limits_x \mathcal{N}(x) f\left(\frac{\mathcal{M}(x)}{\mathcal{N}(x)}\right) dx = E_{\mathcal{N}} \left[f\left(\frac{\mathcal{M}}{\mathcal{N}}\right) \right], \tag{5.13}$$

where $f(.)$ is a convex function and as per the definition, $f(1) = 0$. Using (5.13) and Jensen's inequality, the following relation can be deduced:

$$E_{\mathcal{N}} \left[f\left(\frac{\mathcal{M}}{\mathcal{N}}\right) \right] \geqslant f\left[E_{\mathcal{N}} \left(\frac{\mathcal{M}}{\mathcal{N}}\right) \right]$$

$$\Rightarrow D_f(\mathcal{M} \, || \, \mathcal{N}) \geqslant 0 \quad \left[\text{as } f\left(\int\limits_x \mathcal{M}(x) dx\right) = f(1) = 0\right]. \tag{5.14}$$

The minimum value of $D_f(\mathcal{M} \, || \, \mathcal{N})$ is 0 and is achieved if and only if $\mathcal{M} = \mathcal{N}$. So, $\chi^2_{Pearson} \left[2P_{\mathcal{G}z_c z_s} \, || \, (P_{\mathcal{E}_c \mathcal{E}_s \mathcal{X}} + P_{\mathcal{G}z_c z_s})\right]$ achieves 0 if and only if:

$$2P_{\mathcal{G}z_c z_s}(x, z_c, z_s) = P_{\mathcal{E}_c \mathcal{E}_s \mathcal{X}}(x, z_c, z_s) + P_{\mathcal{G}z_c z_s}(x, z_c, z_s)$$

$$\Rightarrow P_{\mathcal{E}_c \mathcal{E}_s \mathcal{X}}(x, z_c, z_s) = P_{\mathcal{G}z_c z_s}(x, z_c, z_s). \tag{5.15}$$

So, theoretically, joint distribution corresponding to generated/fake encoding becomes identical to the joint distribution corresponding to real encoding. As a result, the generator $\mathcal{G}$ becomes successful to beat the discriminator $\mathcal{D}$, that is, the discriminator cannot determine whether the triplet (x, z_c, z_s) is coming from real joint distribution $P_{\mathcal{E}_c \mathcal{E}_s \mathcal{X}}$ or generated/fake joint distribution $P_{\mathcal{G}z_c z_s}$, and hence, it concludes the proof.

Hence, the adversarial objective term, attributed by the generative module, can be framed as follows:

$$J_{\text{adv}} = J_{\text{G}}(\mathcal{D}) + J_{\text{G}}(\mathcal{G}). \tag{5.16}$$

5.2.3.2 Reconstructive Module

The reconstructive module contains three deep networks: $\mathcal{E}_c$, $\mathcal{E}_s$ and $\mathcal{G}$, as marked in Fig. 5.1 with red dashed line. The reconstructive module also expects the latent color appearance code z_c and latent stain density code z_s to be independent of each other to extract

disentangled color appearance and stain bound information. Now, the joint distribution $P_{\mathcal{G}z_c z_s}$, which is associated with the reconstruction of image samples, can be decomposed as follows:

$$P_{\mathcal{G}z_c z_s}(x, z_c, z_s) = P_{\mathcal{G}}(x \mid z_c, z_s) P_{Z_c Z_s}(z_c, z_s). \tag{5.17}$$

Now, z_c and z_s are assumed to be independent. Hence,

$$P_{\mathcal{G}z_c z_s}(x, z_c, z_s) = P_{\mathcal{G}}(x \mid z_c, z_s) P_{Z_c}(z_c) P_{Z_s}(z_s). \tag{5.18}$$

The inference problem is to compute the joint conditional density of latent codes z_c, z_s given the real data sample x. So,

$$P(z_c, z_s \mid x) = \frac{P_{\mathcal{G}z_c z_s}(x, z_c, z_s)}{P(x)}, \tag{5.19}$$

where $P(x)$ is the marginal likelihood or model evidence, which can be computed as follows:

$$P(x) = \int_{z_c} \int_{z_s} P_{\mathcal{G}z_c z_s}(x, z_c, z_s) dz_c dz_s. \tag{5.20}$$

But, the computation of model evidence is typically intractable. This is due to the fact that the integral in (5.20) requires integration with respect to multi-dimensional latent representations z_c and z_s. So, an auxiliary approximate distribution $Q(z_c, z_s)$, defined in such a way that it has closed form solution, should be used to approximate the inference $P(z_c, z_s \mid x)$ by minimizing the Kullback-Leibler (KL) divergence between $P(z_c, z_s \mid x)$ and $Q(z_c, z_s)$ as follows:

$$D_{KL}[Q(z_c, z_s) \mid\mid P(z_c, z_s \mid x)] = - \int_{z_c} \int_{z_s} Q(z_c, z_s) \log \left[\frac{P(z_c, z_s \mid x)}{Q(z_c, z_s)} \right] dz_c dz_s$$

$$= -E_{Q(z_c, z_s)} \left[\log \left(\frac{P_{\mathcal{G}z_c z_s}(x, z_c, z_s)}{Q(z_c, z_s)} \right) \right] + \log P(x). \tag{5.21}$$

Now, the minimum value of KL divergence, denoted by $D_{KL}[. \mid\mid .]$ is 0. Hence, from (5.21), it can be written that

$$-E_{Q(z_c, z_s)} \left[\log \left(\frac{P_{\mathcal{G}z_c z_s}(x, z_c, z_s)}{Q(z_c, z_s)} \right) \right] + \log P(x) \geqslant 0$$

$$\Rightarrow \log P(x) \geqslant E_{Q(z_c, z_s)} \left[\log \left(\frac{P_{\mathcal{G}z_c z_s}(x, z_c, z_s)}{Q(z_c, z_s)} \right) \right]. \tag{5.22}$$

Here, $\log P(x)$ represents the log evidence and hence, the right hand side of inequality (5.22) represents the evidence lower bound (ELBO). Thus, minimizing the KL divergence in (5.21) is same as maximizing the ELBO in inequality (5.22).

To frame it as minimization objective, taking negative of inequality (5.22), the upper

bound can be obtained as follows:

$$-\log P(x) \leqslant -E_{Q(z_c, z_s)}\left[\log\left(\frac{P_{\mathcal{G}z_c z_s}(x, z_c, z_s)}{Q(z_c, z_s)}\right)\right]. \tag{5.23}$$

Let, the right hand side of inequality (5.23), which is to be minimized, be denoted by F. Then,

$$F = -E_{Q(z_c, z_s)}\left[\log\left(\frac{P_{\mathcal{G}z_c z_s}(x, z_c, z_s)}{Q(z_c, z_s)}\right)\right].$$

Using (5.18), F can be rewritten as follows:

$$F = -E_{Q(z_c, z_s)}\left[\log\left(\frac{P_{\mathcal{G}}(x \mid z_c, z_s)P_{\mathcal{Z}_c}(z_c)P_{\mathcal{Z}_s}(z_s)}{Q(z_c, z_s)}\right)\right]$$

$$= -E_{Q(z_c, z_s)}[\log P_{\mathcal{G}}(x \mid z_c, z_s)] - E_{Q(z_c, z_s)}\left[\log\frac{P_{\mathcal{Z}_c}(z_c)}{Q(z_c, z_s)}\right]$$

$$-E_{Q(z_c, z_s)}\left[\log\frac{P_{\mathcal{Z}_s}(z_s)}{Q(z_c, z_s)}\right] - E_{Q(z_c, z_s)}[\log Q(z_c, z_s)]$$

$$= -E_{Q(z_c, z_s)}[\log P_{\mathcal{G}}(x \mid z_c, z_s)] + D_{KL}[Q(z_c) \mid\mid P_{\mathcal{Z}_c}(z_c)]$$

$$+D_{KL}[Q(z_s) \mid\mid P_{\mathcal{Z}_s}(z_s)] - E_{Q(z_c, z_s)}[\log Q(z_c, z_s)]. \tag{5.24}$$

where the last term in (5.24) represents the entropy defined over $Q(z_c, z_s)$, which acts as a regularizer for auxiliary distribution $Q(z_c, z_s)$. As z_c is assumed to be independent of z_s, from optimization perspective, $D_{KL}[Q(z_c) \mid\mid P_{\mathcal{Z}_c}(z_c)]$ becomes identical to $D_{KL}[Q(z_c, z_s) \mid\mid P_{\mathcal{Z}_c}(z_c)]$. Similarly, optimizing $D_{KL}[Q(z_c, z_s) \mid\mid P_{\mathcal{Z}_s}(z_s)]$ is same as optimizing $D_{KL}[Q(z_s) \mid\mid P_{\mathcal{Z}_s}(z_s)]$. Thus, from (5.24), the reconstruction objective to be minimized can be framed as:

$$J_{\text{rec}} = \underbrace{-E_{Q(z_c, z_s)}[\log P_{\mathcal{G}}(x \mid z_c, z_s)]}_{L_R} - E_{Q(z_c, z_s)}[\log Q(z_c, z_s)]$$

$$+ \underbrace{D_{KL}[Q(z_c) \mid\mid P_{\mathcal{Z}_c}(z_c)]}_{R_1} + \underbrace{D_{KL}[Q(z_s) \mid\mid P_{\mathcal{Z}_s}(z_s)]}_{R_2}, \tag{5.25}$$

where L_R represents the reconstruction loss, whereas R_1 and R_2 denote the regularization terms corresponding to color appearance code z_c and stain density code z_s, respectively.

To solve the minimization problem of (5.25), the priors for latent distributions, corresponding to color appearance code z_c as well as stain density code z_s, are to be assumed in order to compute the KL divergence values. Now, the latent color appearance representation, corresponding to each stained histological image, should capture information corresponding to all the stains involved in the staining routine. Hence, to deal with the overlapping nature of stains, a mixture of probability distributions should be used as the prior for latent multimodal distribution corresponding to color appearance code. But, the

outer tails of a mixture model are prone to outliers and do not contribute adequately in handling overlapping information. Based on this intuition, a mixture of truncated normal distributions, referred to as truncated Gaussian mixture model (tGMM), is used as the prior for latent distribution corresponding to color appearance code. As stain bound information, corresponding to a particular image, is to be kept intact after color normalization, standard normal distribution is used as the prior for latent stain density information. The computation of regularization terms R_1 and R_2 of (5.25) is provided in the appendix.

Definition 5.1. *If X denotes a random variable that follows a truncated normal distribution, then the probability density function (pdf) of X is defined as follows [98]:*

$$\mathcal{F}_X(x \mid \mu, \sigma, \kappa) = \frac{1}{\sigma} \frac{\phi(\frac{x-\mu}{\sigma})}{\Phi(\kappa) - \Phi(-\kappa)} I_{[\mu-\kappa\sigma, \mu+\kappa\sigma]}(x), \tag{5.26}$$

where μ and σ represent the mean and standard deviation, respectively. The truncation factor $\kappa = \frac{b-\mu}{\sigma} = \frac{\mu-a}{\sigma}$ determines the number of standard deviations away from the mean, where a and b denote the left and right truncation points respectively, $\phi(.)$ is the standard normal pdf and $\Phi(.)$ denotes standard normal cumulative distribution function. Here, $I_{[\mu-\kappa\sigma, \mu+\kappa\sigma]}(.)$ represents the indicator function.

Algorithm 5.1 summarizes the training procedure for the proposed method. Once the networks are trained, Algorithm 5.2 is used to generate color normalized source images based on the trained network parameters $\{\theta_{\mathcal{E}_c}, \theta_{\mathcal{E}_s}, \theta_{\mathcal{G}}\}$, obtained via Algorithm 5.1.

5.3 Performance Analysis

This section presents the performance of the proposed deep generative model, termed as TredMiL.

5.3.1 Experimental Setup

A brief description of the algorithms compared, data sets and quantitative indices used, and training details is provided next.

5.3.1.1 Algorithms Compared

The performance of the proposed model is studied extensively and compared with that of

- several state-of-the-art methods for stain estimation and stain separation: plane fitting (PF) [84], HTN due to Li and Plataniotis [132], enhanced PF (EPF) [89], structure-preserving color normalization (SPCN) [22], expectation-maximization (EM) algorithm [134], and rough-fuzzy circular clustering method RFCC$_{vM}$ [105]; and

- several state-of-the-art color normalization algorithms: color transfer (ColTrans) [38], stain color description (SCD) [15], stain normalization using generative adversarial networks (SN-GAN) [41], StainGAN [90], and adversarial stain transfer (AST) [6], as well as PF [84], HTN [132], EPF [89], SPCN [22] and RFCC$_{vM}$ [105].

Algorithm 5.1 Mini-batch stochastic gradient descent training of proposed deep generative model.

Input: Training set images sampled from real data distribution $P_X(x)$, number of training epochs (τ).

Output: Trained parameter set $\{\theta_{\mathcal{D}}, \theta_{\mathcal{E}_c}, \theta_{\mathcal{E}_s}, \theta_{\mathcal{G}}\}$, corresponding to four deep networks $\mathcal{D}$, $\mathcal{E}_c$, $\mathcal{E}_s$ and $\mathcal{G}$, respectively.

1: **for** each epoch $t = 0$ to $\tau - 1$ **do**

- Sample mini-batch of M images $\{x_m\}_{m=1}^M$ from real data distribution $P_X(x)$.

- Generate M color appearance codes $\{z_{c_m}\}_{m=1}^M$ by using $\mathcal{E}_c(x_m; \theta_{\mathcal{E}_c})$.

- Generate M stain density codes $\{z_{s_m}\}_{m=1}^M$ by using $\mathcal{E}_s(x_m; \theta_{\mathcal{E}_s})$.

- Generate reconstructed images $\{\hat{x}_m\}_{m=1}^M$ by using $\mathcal{G}(z_{c_m}, z_{s_m}; \theta_{\mathcal{G}})$.

- Sample mini-batch of M color appearance codes $\{\acute{z}_{c_m}\}_{m=1}^M$ from latent color appearance prior $P_{\mathcal{Z}_c}(z_c)$.

- Sample mini-batch of M stain density codes $\{\acute{z}_{s_m}\}_{m=1}^M$ from latent stain density prior $P_{\mathcal{Z}_s}(z_s)$.

- Update discriminator $\mathcal{D}$ by descending in the direction of stochastic gradient:

$$\nabla_{\theta_{\mathcal{D}}} \left(\frac{1}{M} \sum_{m=1}^M \left[(A - \mathcal{D}(x_m, z_{c_m}, z_{s_m}; \theta_{\mathcal{D}}))^2 + (B - \mathcal{D}(\hat{x}_m, \acute{z}_{c_m}, \acute{z}_{s_m}; \theta_{\mathcal{D}}))^2 \right] \right)$$

- Update color appearance encoder $\mathcal{E}_c$ by descending in the direction of stochastic gradient:

$$\nabla_{\theta_{\mathcal{E}_c}} \left(\frac{1}{M} \sum_{m=1}^M \left[(A - \mathcal{D}(x_m, z_{c_m}, z_{s_m}; \theta_{\mathcal{D}}))^2 + D_{KL}(z_{c_m} \| \acute{z}_{c_m}) \right] \right)$$

- Update stain density encoder $\mathcal{E}_s$ by descending in the direction of stochastic gradient:

$$\nabla_{\theta_{\mathcal{E}_s}} \left(\frac{1}{M} \sum_{m=1}^M \left[(A - \mathcal{D}(x_m, z_{c_m}, z_{s_m}; \theta_{\mathcal{D}}))^2 + D_{KL}(z_{s_m} \| \acute{z}_{s_m}) \right] \right)$$

- Update decoder/generator $\mathcal{G}$ by descending in the direction of stochastic gradient:

$$\nabla_{\theta_{\mathcal{G}}} \left(\frac{1}{M} \sum_{m=1}^M \left[(C - \mathcal{D}(\hat{x}_m, \acute{z}_{c_m}, \acute{z}_{s_m}; \theta_{\mathcal{D}}))^2 - \log(\hat{x}_m) \right] \right)$$

2: **end for**

3: Stop.

Algorithm 5.2 Color normalization of source images via color appearance mapping.

Input: A set of N source images $\{x_n^S\}_{n=1}^N$, template image x^T, trained network parameters $\{\theta_{\mathcal{E}_c}, \theta_{\mathcal{E}_s}, \theta_{\mathcal{G}}\}$ corresponding to deep networks $\mathcal{E}_c$, $\mathcal{E}_s$ and $\mathcal{G}$, obtained via Algorithm 4.1.

Output: Color normalized source images $\{\tilde{x}_n^S\}_{n=1}^N$.

1: Generate latent color appearance code z_c^T and latent stain density code z_s^T corresponding to template image x^T by using $\mathcal{E}_c(x^T; \theta_{\mathcal{E}_c})$ and $\mathcal{E}_s(x^T; \theta_{\mathcal{E}_s})$, respectively.

2: **for** each image in source image set $\{x_n^S\}_{n=1}^N$ **do**

- Generate latent color appearance code $z_{c_n}^S$ and latent density code $z_{s_n}^S$ by using $\mathcal{E}_c(x_n^S; \theta_{\mathcal{E}_c})$ and $\mathcal{E}_s(x_n^S; \theta_{\mathcal{E}_s})$, respectively.

- Feed template image latent color appearance code z_c^T and source image latent stain density code $z_{s_n}^S$ to the generator $\mathcal{G}$ and by using $\mathcal{G}(z_c^T, z_{s_n}^S; \theta_{\mathcal{G}})$ generate normalized source image $\tilde{x}_n^S$

3: **end for**

4: Stop.

5.3.1.2 Data Sets Used

To analyze the performance of different variants of the proposed model and several existing algorithms in stain estimation, stain separation and color normalization, two publicly available data sets are used in this study:

- UCSB Data: This breast cancer cell data set was published by the University of California, Santa Barbara [36] for the task of diseased tissue classification. This H&E stained data set contains a total of 58 images: 32 non-cancerous benign cell and 26 cancerous malignant cell images. The UCSB data comprises of 10 biopsy sets: (9×6) images $+ (1 \times 4)$ images. Each image in this data set contains associated ground-truth binary mask image for nuclei segmentation. Each image has a resolution of 896×768 and is stored in 24 bit non-linear RGB format.

- CMU Data: This H&E stained histology image set was published by the bimagicLab in Carnegie Mellon University [89]. Though this data set contains only 3 images, it possesses high importance as each image contains ground-truth stain decomposition images, that is, separate H-stained and E-stained images corresponding to each histological image. Each image has a resolution of 1280×1024 and is stored in 48 bit linear RGB format.

In order to establish the importance of color normalization in histological image segmentation, following nuclei segmentation data set is also used in this study:

- TCGA Data: This H&E stained tissue image data set [97] consists of 30 whole slide images (WSIs) corresponding to several organs, such as, kidney, lung, liver, prostate and so on. The WSIs were downloaded from The Cancer Genomic Atlas (TCGA). Since processing WSIs is computationally very expensive, sub-images of size 1000×1000 were cropped from nuclei-dense regions. This data set contains more than 21,000 annotated nuclei.

5.3.1.3 Quantitative Indices

To analyze the performance of different variants of the TredMiL model and several existing algorithms in stain estimation, element-wise standard deviation is considered. The images within a particular biopsy set are generally collected from the same pathology lab. So, they are expected to go through similar staining routine and identical storage condition, and thus they should exhibit consistent stain vectors. The performance analysis of stain vector estimation with respect to element-wise standard deviation (σ) was originally reported in [132]. In this study, the standard deviation values corresponding to R, G and B channels for all biopsy sets are presented through heatmap representation, where low and high σ values are represented by deep blue shade and deep red shade, respectively. So, a good stain vector estimation method should exhibit color shade close to deep blue to ensure the consistency of the estimated stain vectors within a particular biopsy set. The σ values in all heatmap representations are reported in 10^{-2} scale.

On the other hand, the performance of different color normalization methods is evaluated using normalized median intensity (NMI) [74], between-image color constancy (BiCC) index and within-set color constancy (WsCC) index [105]. The most important aspect of color normalization is to minimize both within-image and between-image color variation among the images of a single biopsy set with respect to a particular region of interest (ROI). The UCSB data contains binary ground-truth images for nuclei segmentation corresponding to each histological image. So, the nuclei regions are considered to be the ROIs in this study. Though NMI [74] evaluates quantitatively the color consistency of a specific ROI within a single image, it cannot capture the color consistency of the ROI between the images of the same biopsy set. In order to deal with the above problem, two other quantitative indices, namely, BiCC and WsCC, have been introduced in [105] to evaluate the performance of different methods in reducing the variation in color distribution of specific ROI at both image-level and biopsy-level, respectively. The performance of different variants of the proposed TredMiL model and several color normalization methods is evaluated based on NMI, BiCC and WsCC, using box and whisker plots, where the top and bottom boundary bars of each box represent the upper and lower quartiles, respectively, central line denotes the median, whiskers are extended upto three standard deviations from the mean, and the outliers are denoted by '+'.

5.3.1.4 Training Details

The value of truncation factor κ is varied from 1.0 to 3.0 with an increment of 0.5, while the value of relative importance of adversarial term λ_{adv} is varied within the range of 0.1 to 1.0, with an increment of 0.1. To find out the optimum values of κ and λ_{adv}, corresponding to a particular image, the NMI is used. The model is trained in leave-one-biopsy-out manner, that is, the model is trained with overlapping image patches of size 256×256, corresponding to the images of $(N-1)$ sets out of total N biopsy sets, and the analysis is performed on the non-overlapping 256×256 size image patches of remaining one biopsy set. The image patches of one biopsy set are considered to be the test set images and the remaining $(N-1)$ biopsy set image patches are used as training set images. This process is repeated for each individual biopsy set independently. The validation set is formed by selecting 10% image patches randomly from the training set. The incorporation of validation set prevents the model from over-fitting and ensures early stopping based on validation loss.

In the current study, the proposed TredMiL model is trained with 7600 overlapping image patches of UCSB data set, while the validation set is comprised of 720 overlapping patches, and the test set contains 696 non-overlapping image patches.

5.3.2 Effectiveness of the Proposed Model

The effectiveness of the proposed TredMiL model is established through the following analysis.

5.3.2.1 Ablation Study
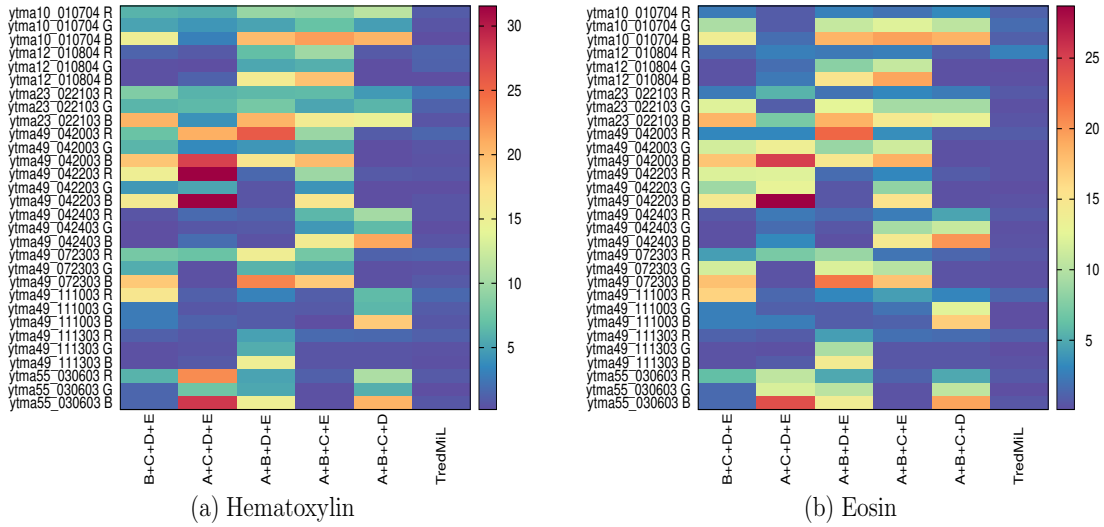
Figure 5.2: Performance analysis of stain vector estimation of the proposed TredMiL model and all five combinations of constituent terms of objective function.

The objective function of the proposed TredMiL model contains five terms: the first term is basically the adversarial term J_{adv} and the last four terms come from the reconstruction term J_{rec}. Let, these five terms be denoted as: $A = J_{\text{adv}}$, $B = -E_{Q(z_c,z_s)}[\log P_{\mathcal{G}}(x \mid z_c, z_s)]$, $C = D_{KL}[Q(z_c) \parallel P_{\mathcal{Z}_c}(z_c)]$, $D = D_{KL}[Q(z_s) \parallel P_{\mathcal{Z}_s}(z_s)]$ and $E = -E_{Q(z_c,z_s)}[\log Q(z_c, z_s)]$. The significance of each individual term is established through an ablation study, where each term is omitted from the objective function and the performance of the model in stain vector estimation is observed in the absence of that particular term. From the heatmap representation presented in Fig. 5.2(a), it is evident that the proposed TredMiL model with all five constituent terms performs better than every individual combination in the estimation of H-stain vector. Similarly, the heatmap presented in Fig. 5.2(b) ensures that the TredMiL model outperforms every individual combination in case of E-stain vector estimation.

The performance of the proposed TredMiL model in color normalization is also compared with that of other five combinations of the constituent terms of the objective function with respect to NMI, BiCC and WsCC, and the corresponding results are reported in Fig. 5.3. From the boxplot representation presented in Fig. 5.3, it can be observed that
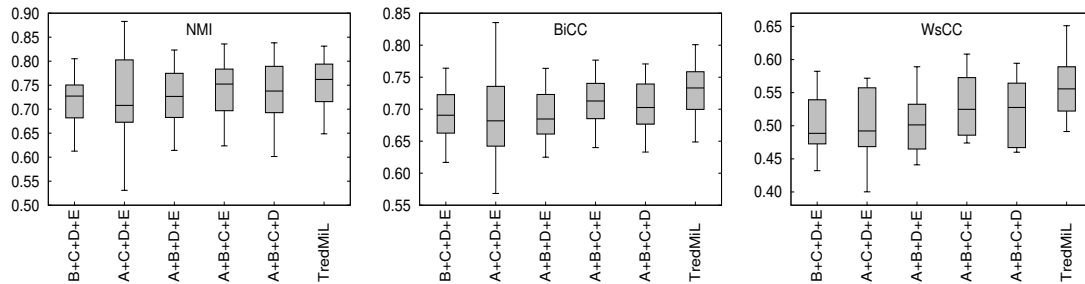
Figure 5.3: Performance analysis of color normalization of the proposed TredMiL model and all five combinations of constituent terms of objective function.

TredMiL attains highest median with respect to all three quantitative indices used. The statistical significance analysis of the proposed TredMiL model is performed with respect to p-values computed through Wilcoxon signed-rank (one-tailed) and paired-t (one-tailed) tests. From the p-values reported in Table 5.1, it can be noticed that the proposed TredMiL model performs significantly better than every individual combination of the constituent terms of the objective function, considering 95% confidence level.

5.3.2.2 Independence Between z_c and z_s
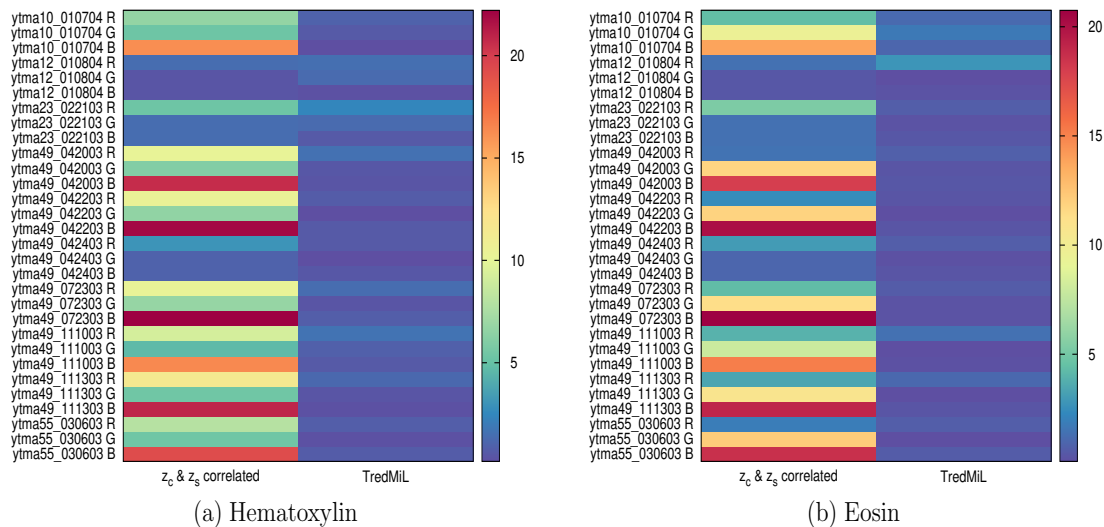
Figure 5.4: Performance analysis of stain vector estimation of proposed TredMiL model and the variant where color appearance code z_c and stain density information z_s are correlated.

The proposed TredMiL model assumes that the latent color appearance information z_c and latent stain density information z_s are independent of each other. The performance of the proposed model in stain vector estimation is compared with that of the model where z_c and z_s are assumed to be correlated, and the corresponding results are reported in Fig. 5.4. It can be observed from the heatmap representation presented in Fig. 5.4(a) that in case of H-stain, the proposed model outperforms the model where z_c and z_s are assumed to be

Table 5.1: Statistical Significance Analysis of Different Algorithms. The p-values, marked in italics, indicate that the proposed method is better than the existing one, but not significantly.

Different Methods	NMI		BiCC		WsCC	
	Wilcoxon	Paired-t	Wilcoxon	Paired-t	Wilcoxon	Paired-t
B+C+D+E	3.45E-11	3.88E-15	1.75E-11	1.30E-19	2.53E-03	3.25E-05
A+C+D+E	5.87E-03	4.06E-03	6.65E-05	1.91E-05	3.72E-02	3.10E-02
A+B+D+E	4.24E-10	1.88E-11	1.75E-11	1.12E-17	2.53E-03	1.27E-04
A+B+C+E	2.35E-08	3.90E-09	1.75E-11	8.93E-16	2.53E-03	9.53E-05
A+B+C+D	5.16E-05	1.44E-05	2.24E-10	4.19E-10	3.46E-03	4.17E-03
z_c&z_s correlated	4.69E-08	2.20E-08	2.80E-11	5.55E-13	2.53E-03	4.42E-04
C-L+D-R	3.87E-08	1.13E-09	5.47E-11	4.07E-17	2.53E-03	2.34E-05
C-L+D-L	8.13E-11	9.11E-14	2.95E-11	2.57E-17	2.53E-03	2.02E-04
C-R+D-R	5.70E-08	4.30E-07	1.50E-10	8.48E-10	2.53E-03	6.40E-03
GMM+NSGAN	7.47E-08	2.50E-08	1.85E-11	4.26E-16	2.53E-03	1.70E-04
GMM+LSGAN	4.67E-04	*2.74E-01*	3.51E-05	*1.76E-01*	3.72E-02	*2.71E-01*
tGMM+NSGAN	6.14E-04	*2.70E-01*	5.00E-05	*1.47E-01*	4.63E-02	*2.88E-01*
ColTrans	1.75E-11	2.16E-18	1.75E-11	2.75E-32	2.53E-03	1.40E-08
PF	7.82E-11	9.05E-18	1.75E-11	2.87E-27	2.53E-03	1.57E-07
EPF	4.46E-11	3.79E-18	1.75E-11	1.83E-27	2.53E-03	4.63E-08
SCD	1.29E-10	1.85E-16	2.16E-11	9.67E-21	2.53E-03	6.31E-05
HTN	5.92E-07	3.85E-08	2.15E-09	1.02E-12	3.46E-03	5.75E-04
SPCN	2.05E-11	1.60E-20	1.75E-11	2.02E-28	2.53E-03	3.61E-07
SN-GAN	1.11E-08	2.70E-11	1.85E-11	3.15E-22	2.53E-03	6.03E-06
StainGAN	1.75E-11	6.10E-23	1.75E-11	4.64E-35	2.53E-03	2.36E-09
AST	9.57E-11	8.78E-16	1.75E-11	7.13E-27	2.53E-03	1.21E-06
RFCC$_{vM}$	3.89E-06	7.57E-07	2.70E-06	4.28E-07	6.26E-03	3.63E-03

correlated. Similarly, from Fig. 5.4(b), it is evident that TredMiL performs better than the model with correlated z_c and z_s in case of E-stain.

To evaluate the performance of the proposed model in color normalization, its performance is compared with that of the model where z_c and z_s are assumed to be correlated. From the boxplot representation provided in Fig. 5.5, it is evident that the proposed Tred-MiL model with independent z_c and z_s attains higher median with respect to NMI, BiCC and WsCC than the model where z_c and z_s are assumed to be correlated. As the model with correlated z_c and z_s loses significant amount of structural information during mapping, the proposed model with independent z_c and z_s performs better than the correlated model. From the p-values reported in Table 5.1, it is evident that the proposed TredMiL model performs significantly better than the model where z_c and z_s are assumed to be correlated.

5.3.2.3 Activation Function

In the proposed TredMiL model, LeakyReLU and ReLU are used as the activation functions for the color appearance encoder $\mathcal{E}_c$ and the discriminator $\mathcal{D}$, respectively. In order to establish the importance of the proposed TredMiL model, the performance of the proposed model in stain vector estimation is compared with that of the other three combinations
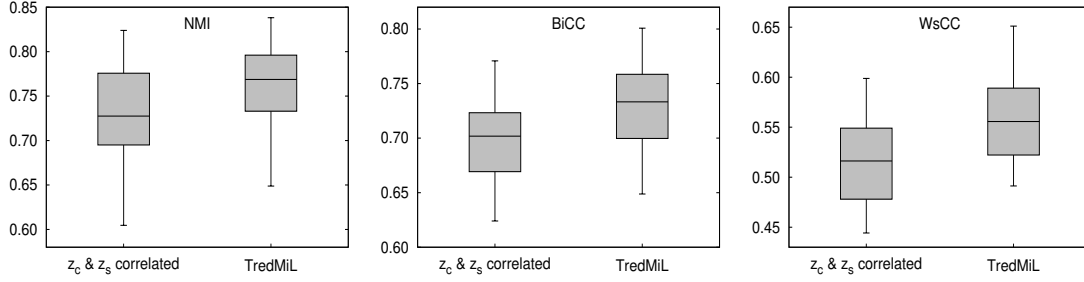
Figure 5.5: Performance analysis of color normalization of the proposed TredMiL model and the variant where color appearance code z_c and stain density information z_s are correlated.

(a) Hematoxylin

(b) Eosin
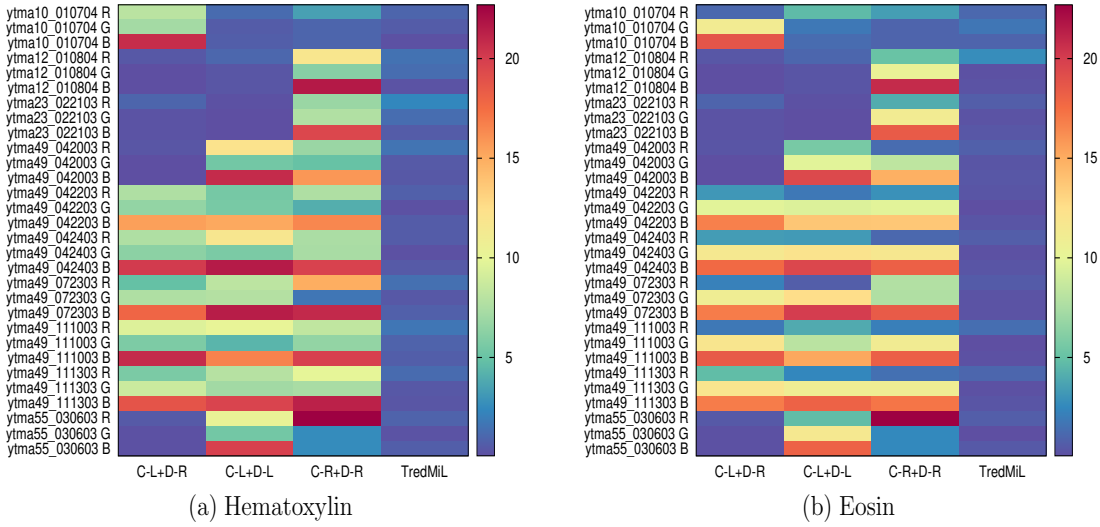
Figure 5.6: Comparative performance analysis of stain vector estimation of the proposed TredMiL model and different combinations of color appearance encoder $\mathcal{E}_c$ and discriminator $\mathcal{D}$ activation functions.

of $\mathcal{E}_c$ and $\mathcal{D}$ activations, and the results are presented in Fig. 5.6. From the heatmap presented in Fig. 5.6(a), it is evident that the TredMiL model performs better than other three combinations in case of H-stain. Similarly, it can be observed from Fig. 5.6(b) that the proposed model outperforms the other three combinations of $\mathcal{E}_c$ and $\mathcal{D}$ activations corresponding to E-stain.

The performance of the proposed TredMiL model in color normalization is also compared with that of other three combinations of $\mathcal{E}_c$ and $\mathcal{D}$ activations, and the associated results are reported in Fig. 5.7. From the results reported in Fig. 5.7, it can be noted that the proposed model performs better than other three combinations of $\mathcal{E}_c$ and $\mathcal{D}$ activations with respect NMI, BiCC and WsCC indices. It can also be observed from the p-values reported in Table 5.1 that the proposed TredMiL model performs significantly better than these three combinations. The proposed model performs better than other three combinations due to following two reasons: (a) ReLU as $\mathcal{E}_c$ activation function handles the vanishing gradient problem, which helps in updating the network parameters properly, and
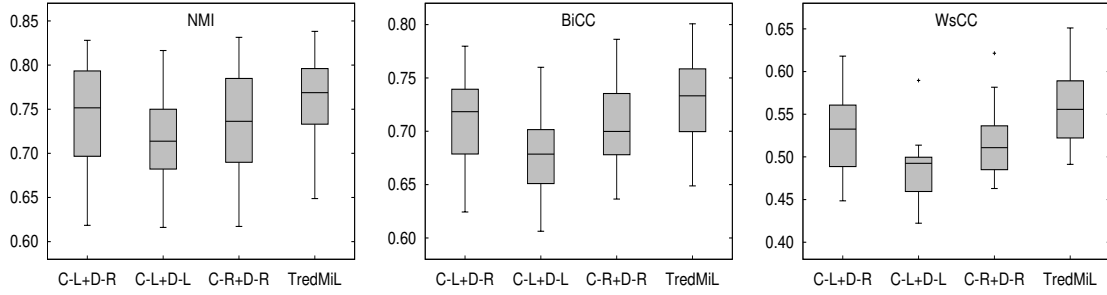
Figure 5.7: Comparative performance analysis of color normalization of the proposed Tred-MiL model and different combinations of color appearance encoder $\mathcal{E}_c$ and discriminator $\mathcal{D}$ activation functions.

(b) LeakyReLU as $\mathcal{D}$ activation handles the negative values, which helps the model to fit or train from the data.

5.3.2.4 Mixture Model and GAN Formulation

(a) Hematoxylin

(b) Eosin
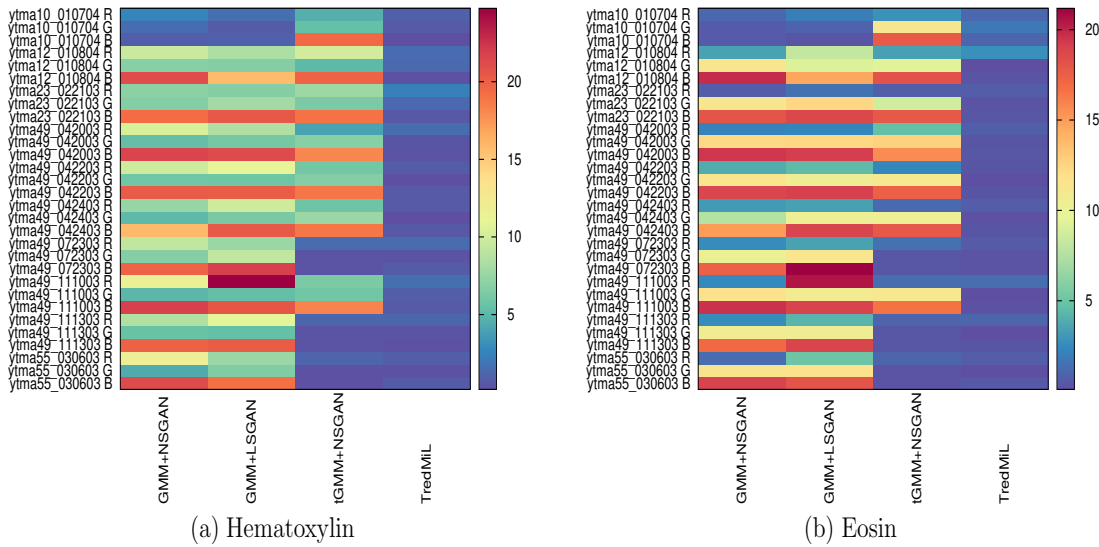
Figure 5.8: Comparative performance analysis of stain vector estimation of the proposed TredMiL model, and mixture model and GAN formulation based other variants.

The proposed TredMiL model utilizes LSGAN instead of non-saturating GAN (NS-GAN). The performance of TredMiL model in stain vector estimation is compared with that of mixture model and GAN formulation based different variants, and the results are presented in Fig. 5.8. The results presented through heatmap in Fig. 5.8(a) and Fig. 5.8(b) exhibit the fact that the TredMiL model outperforms other three combinations in estimating H-stain and E-stain, respectively.

The performance of the proposed TredMiL model with respect to color normalization is also compared with that of other three mixture models and GAN formulation based
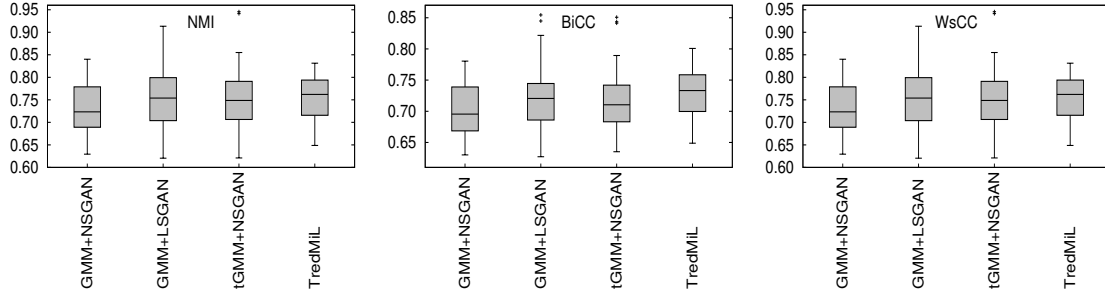
Figure 5.9: Performance analysis of color normalization of the proposed TredMiL model, and mixture model and GAN formulation based other variants.

combinations. It can be observed from Fig. 5.9 that though different combinations of mixture models and GAN formulations exhibit comparable results, TredMiL achieves highest median values with respect to all three indices. The proposed model performs better than other GMM based formulations as it can reduce the effect of outliers. Also, the TredMiL model outperforms the NSGAN based formulations as the convexity property of least square function prevents vanishing gradient problem and negates unstable behavior of alternating min-max optimization. From the p-values reported in Table 5.1, it can be observed that the proposed TredMiL model performs significantly better than the combination of GMM and NSGAN. It can also be noted from Table 5.1 that the proposed TredMiL model performs better, but not significantly, than remaining two combinations, namely, GMM + LSGAN, and tGMM + NSGAN, with respect to paired-t test.

5.3.2.5 Significance in Histological Image Segmentation

(a) Original   (b) Ground-truth   (c) Non-normalized   (d) Normalized
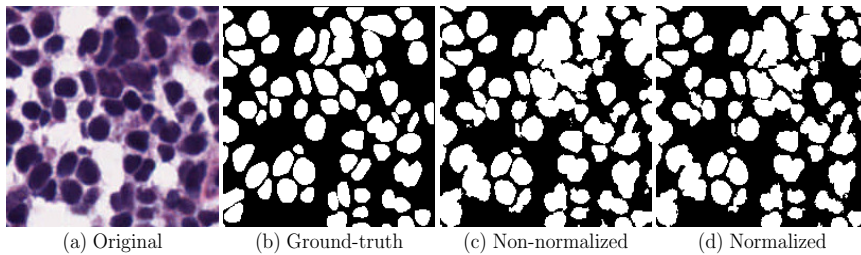
Figure 5.10: Performance analysis of Watershed algorithm in nuclei segmentation from histological images before and after color normalization: (a) Original histological image, (b) Ground-truth; segmentation performance- (c) before color normalization (non-normalized), (d) after normalization (normalized).

The color normalization is considered to be an important pre-processing step for histological image segmentation. To examine the significance of color normalization in nuclei segmentation, the images from TCGA data set [97] are used to analyze the performance of nuclei segmentation before and after color normalization. The Watershed segmentation method is used in this study. The performance of nuclei segmentation on one image patch extracted from one of the images of the TCGA data set is provided in Fig. 5.10. From the

qualitative result presented in Fig. 5.10, it is clear that the Watershed segmentation method performs better on the image patch after color normalization than the image patch before normalization. The significance of color normalization has also been validated through the average Dice co-efficient values reported in Table 5.2. After color normalization, the mean and median values of Dice co-efficient are increased by 3.38% and 3.78%, respectively. The results presented in Fig. 5.10 and Table 5.2 highlight the effectiveness of color normalization of histological images.

Table 5.2: Performance analysis of nuclei segmentation before and after color normalization through Dice co-efficient. Mean, median and standard deviation of Dice co-efficient are computed considering 30 images of TCGA data.

	Mean	Median	Stddev
Before normalization	0.681253	0.684582	0.103128
After normalization	**0.704252**	**0.710469**	**0.097357**

5.3.2.6 Color Variation Due to Fade Color Effect

(a) Opacity=10   (b) Opacity=25   (c) Opacity=50   (d) Opacity=75   (e) Opacity=90
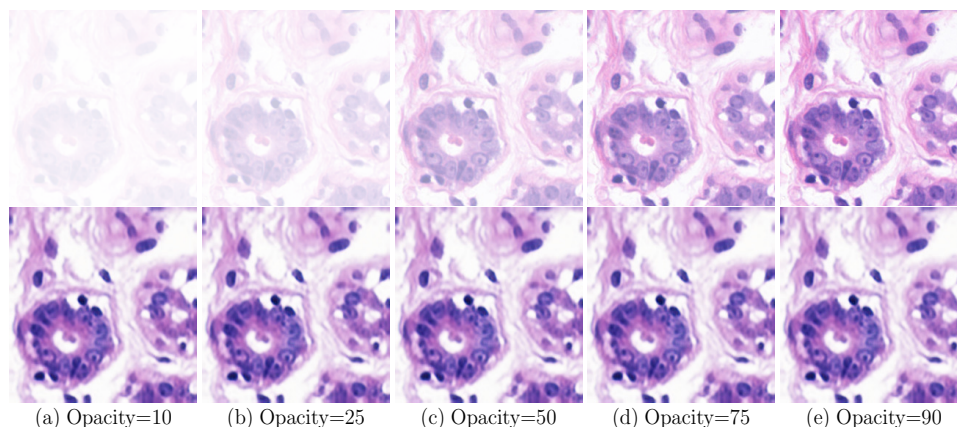
Figure 5.11: Performance analysis of color normalization on images with different opacity levels. Opacity $= x$ means the fade percentage is $(100 - x)$. Top-row: histological images with different opacity levels (10, 25, 50, 75, 90); bottom-row: color normalized images corresponding to top-row images.

The performance of the proposed TredMiL model has also been analyzed for the images which suffer from variation in color representation due to the fade color effect. The images of the UCSB data set are faded manually and the trained model is used to generate corresponding normalized images. The images are faded based on different opacity levels: 10, 25, 50, 75 and 90. Here, opacity $= x$ means fade percentage is $(100 - x)$.

The qualitative analysis of the color normalization performance is exhibited in Fig. 5.11. The color normalization performance is analyzed with respect to NMI, BiCC and WsCC indices, and the associated results are reported in Table 5.3. Table 5.3 also reports the performance of image reconstruction through root mean square error (RMSE) measure. The performance of the model increases with the increase in opacity levels. Though there

Table 5.3: NMI, BiCC, WsCC and RMSE for different opacity values. Higher the opacity means lower the percentage of fade.

Measures	Opacity=10	Opacity=25	Opacity=50	Opacity=75	Opacity=90
			NMI		
Mean	0.745462	0.746109	0.746287	0.746828	**0.746865**
Median	0.749269	0.749269	0.750223	0.750160	**0.750832**
Stddev	0.064107	0.063932	0.064185	0.064862	**0.063708**
			BiCC		
Mean	0.713808	0.714839	0.715560	0.715690	**0.716603**
Median	0.711579	0.711440	0.712393	0.712125	**0.717622**
Stddev	0.053526	0.052893	**0.052457**	0.052589	0.053250
			WsCC		
Mean	0.537290	0.537419	0.538261	0.538496	**0.539655**
Median	0.529401	0.530828	0.530629	0.530824	**0.532263**
Stddev	0.070444	0.070676	**0.069820**	0.070128	0.070296
			RMSE		
Mean	4.747402	2.271538	1.887777	1.757888	**1.731400**
Median	3.408091	1.975749	1.762917	1.699476	**1.671057**
Stddev	3.139788	0.942396	0.546057	0.387382	**0.357084**

is a substantial loss in color appearance information in case of low opacity values, the TredMiL model can normalize the images and reduces the effect of fading. This is possible because of the fact that the proposed model can capture structural information through stain density information, which is independent of the color appearance information, and the mapping is performed based on the color appearance information extracted from the template image.

5.3.2.7 Impact of Dust Artifacts

(a) Original   (b) Dust film   (c) Dust added   (d) Normalized
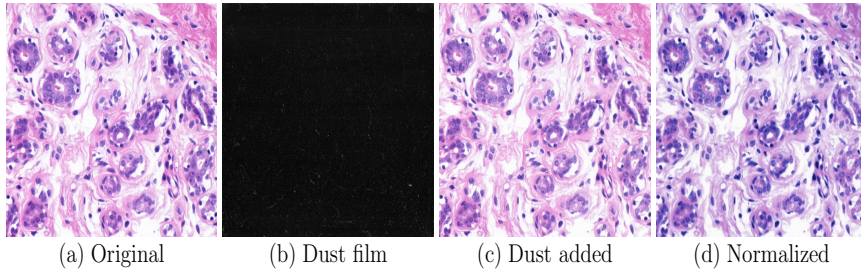
Figure 5.12: Analysis of color normalization performance on UCSB image in the presence of dust particles.

The performance of TredMiL model is also analyzed in the presence of dust artifacts. The dust particles are added to the UCSB images by using dust film images. The qualitative performance of color normalization is presented in Fig. 5.12. From the results reported in Table 5.4, it can be observed that the TredMiL performs considerably well in the presence of dust artifacts. This is due to the fact that the tGMM is able to capture the mixtures of color appearance information of H-stain and E-stain, and the truncation of outer tails of

Table 5.4: Analysis of color normalization performance on dust-added UCSB images in terms of mean, median and standard deviation of NMI, BiCC, and WsCC indices.

	Mean	Median	Stddev
NMI	0.727792	0.729877	0.060089
BiCC	0.703538	0.699569	0.050127
WsCC	0.569785	0.564891	0.064487

tGMM can ignore the color distribution corresponding to outliers as well as dust particles.

5.3.3 Comparison with Existing Methods

Finally, the performance of the proposed TredMiL model is compared with that of several state-of-the-art algorithms.

5.3.3.1 Stain Vector Estimation

(a) Hematoxylin

(b) Eosin
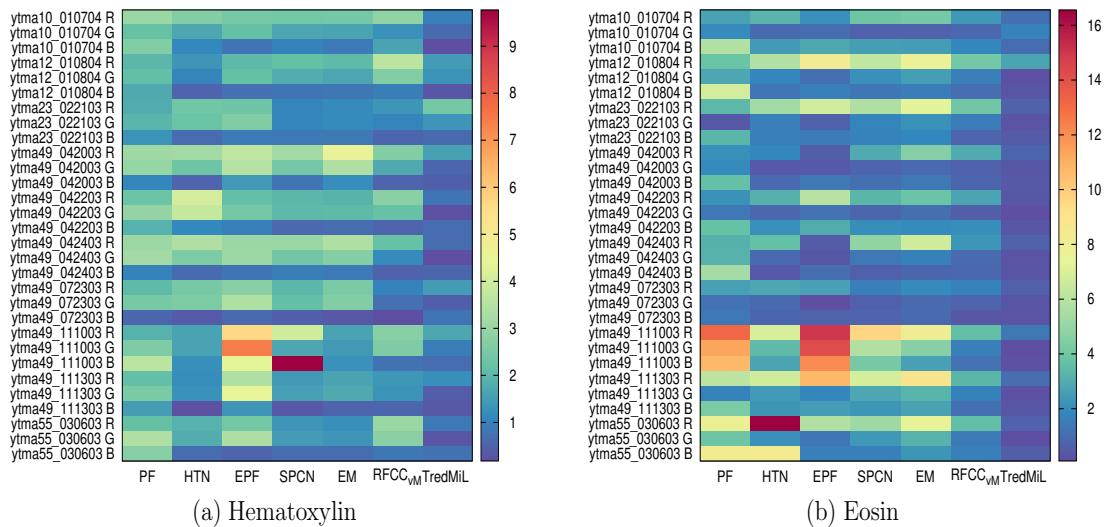
Figure 5.13: Comparative performance analysis of stain vector estimation of the proposed TredMiL model and (a)-(b) different existing methods such as PF, HTN, EPF, SPCN, EM and $RFCC_{vM}$ in stain vector estimation (H-stain and E-stain, respectively).

The performance of the proposed TredMiL model in stain vector estimation is compared with that of different state-of-the-art methods, and the results, corresponding to H-stain and E-stain, are reported via heatmap in Fig. 5.13(a) and Fig. 5.13(b), respectively. From the heatmap presented in Fig. 5.13(a), it can be observed that TredMiL performs better than state-of-the-art stain vector estimation methods in case of H-stain. From Fig. 5.13(b), it is also clearly visible that TredMiL achieves lowest σ values in most of the cases compared to the existing methods in the estimation of E-stain representative vector.
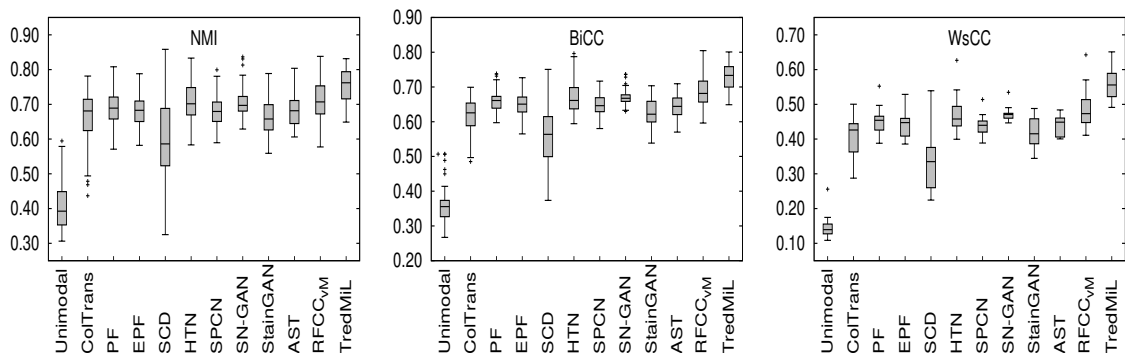
Figure 5.14: Comparative performance analysis of color normalization of the proposed TredMiL model and state-of-the-art methods such as ColTrans, PF, EPF, SCD, HTN, SPCN, SN-GAN, StainGAN, AST and RFCC$_{vM}$ in color normalization.

Table 5.5: Comparison of structural similarity among different existing methods. Higher structural similarity value means structure preservation is better after color normalization.

Methods	Mean	Median	Stddev
ColTrans	0.831310	0.840256	0.036779
PF	0.969343	0.972036	0.023265
EPF	0.987143	0.994131	0.016391
SCD	0.682923	0.682006	0.059083
HTN	0.986894	0.997557	0.035390
SPCN	0.964949	0.972214	0.029760
SN-GAN	0.719060	0.725737	0.065482
StainGAN	0.778872	0.779981	0.035009
AST	0.800084	0.803421	0.061924
RFCC$_{vM}$	0.988245	0.994146	0.012432
TredMiL	**0.997830**	**0.997978**	**0.000782**

5.3.3.2   Color Normalization

The results presented in Fig. 5.14 depict the fact that the proposed model exhibits highest median values than other existing methods with respect to all three quantitative indices, namely, NMI, BiCC and WsCC. Analyzing the p-values reported in Table 5.1, it can be seen that the proposed method performs significantly better than the state-of-the-art color normalization methods. The qualitative performances of different color normalization algorithms are presented in Fig. 5.15, Fig. 5.16 and Fig. 5.17. The results reported in Table 5.1 and Fig. 5.14- 5.17 establish the fact that the proposed method outperforms state-of-the-art color normalization algorithms as per as color consistency after normalization is concerned.

5.3.3.3   Stain Separation

The most important aspect of CMU data set is that each H&E stained image contains individual stain decomposition ground-truth H-stain and E-stain images. As a result, KL divergence and signal-to-noise ratio (SNR) are used to analyze the performance of the
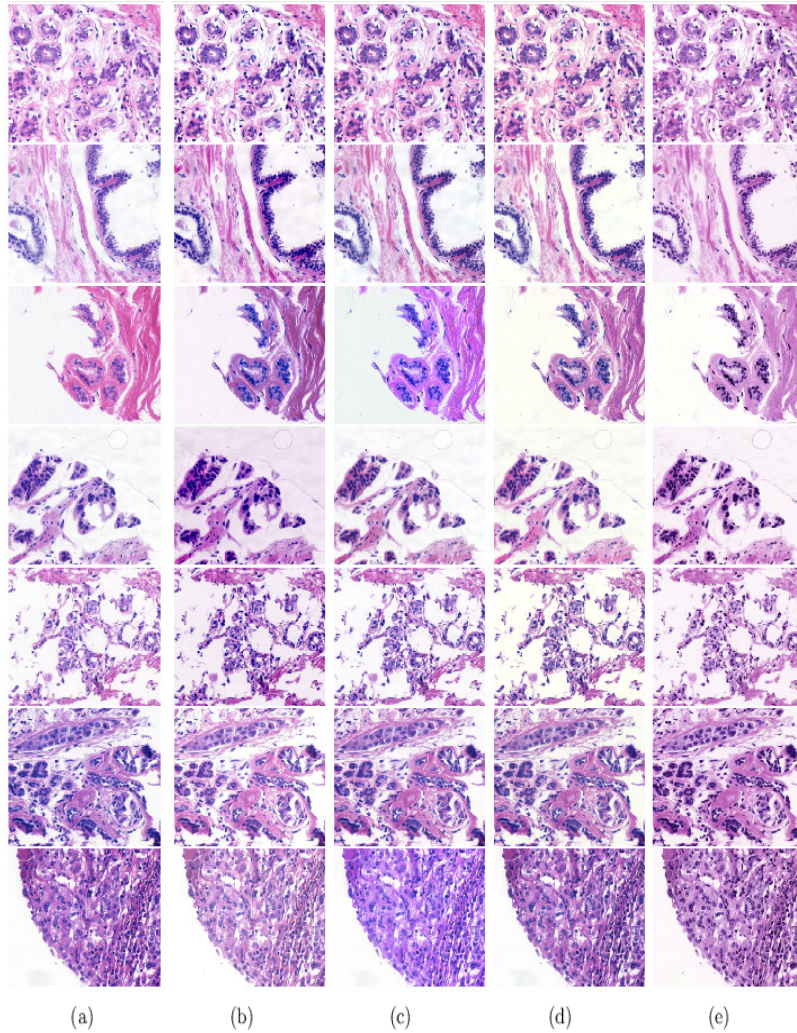
Figure 5.15: (a) Original images of UCSB data set; and color normalized images obtained using different state-of-the-art color normalization methods: (b) ColTrans, (c) PF, (d) EPF and (e) SCD.

proposed TredMiL model in stain separation. The generalization ability of the TredMiL model can be highlighted by the fact that the model trained on UCSB data can be applied on other data such as CMU data. This is possible because both UCSB and CMU images are H&E stained histology images, and the trained model assumes the color appearance information and stain density information to be independent of each other. The performance of the TredMiL model in stain separation is compared with that of other existing stain separation algorithms, and the results, corresponding to KL divergence and SNR, are presented in Fig. 5.18 through the mean value of each quantitative index along with the error bars. Considering three H&E stained histological images of CMU data, the mean, minimum and maximum values, corresponding to each quantitative index, are computed for each individual stain. The results reported in Fig. 5.18 ensure that the proposed model attains lower mean for symmetric KL divergence and higher mean for SNR, compared to

91
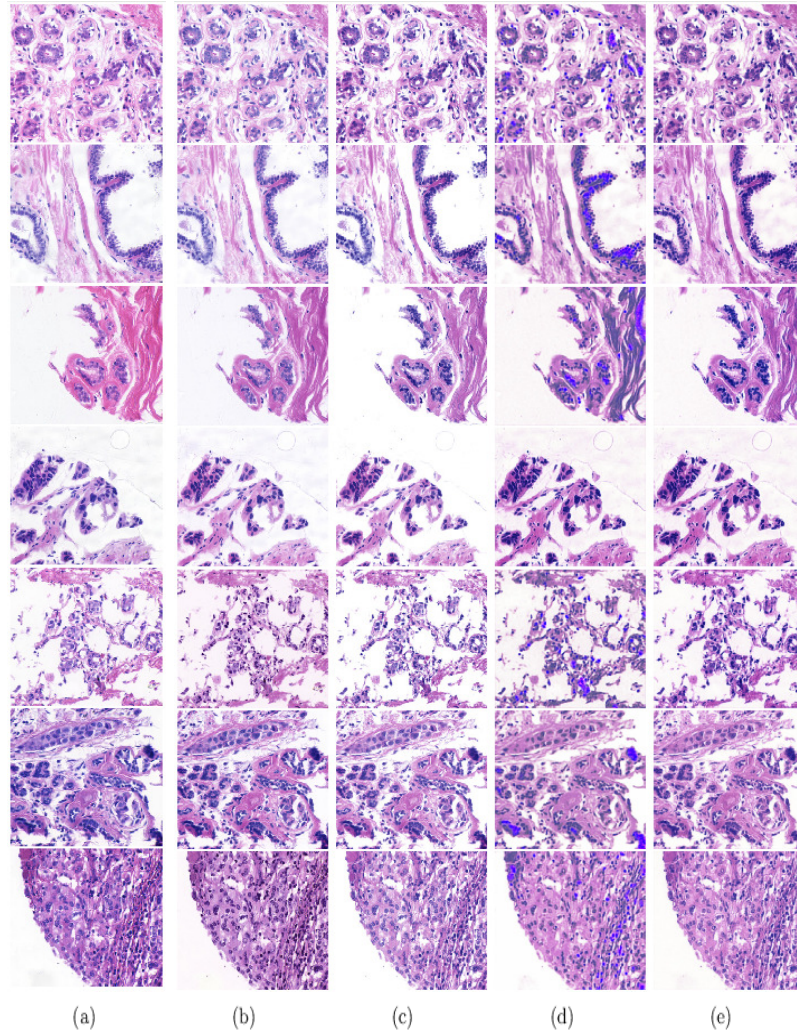
Figure 5.16: (a) Original images of UCSB data set; and color normalized images obtained using different state-of-the-art color normalization methods: (b) HTN, (c) SPCN, (d) SN-GAN and (e) StainGAN.

each of the existing stain separation methods. The qualitative performance analysis of the proposed and different existing methods is provided in Fig. 5.19. It can be observed from Fig. 5.19 that only the proposed method, RFCC$_{vM}$ and EPF are capable of extracting the inherent structures of biological components, highlighted by both H-stain and E-stain.

5.3.3.4 Preservation of Structural Similarity

The aim of color normalization is to reduce the color variation present among images without hampering the structural information contained in the images. The preservation of structural information of different existing color normalization algorithms is analyzed with respect to a structural similarity measurement index, called structural similarity index measure (SSIM) [149]. If the structural similarity between each of the non-normalized
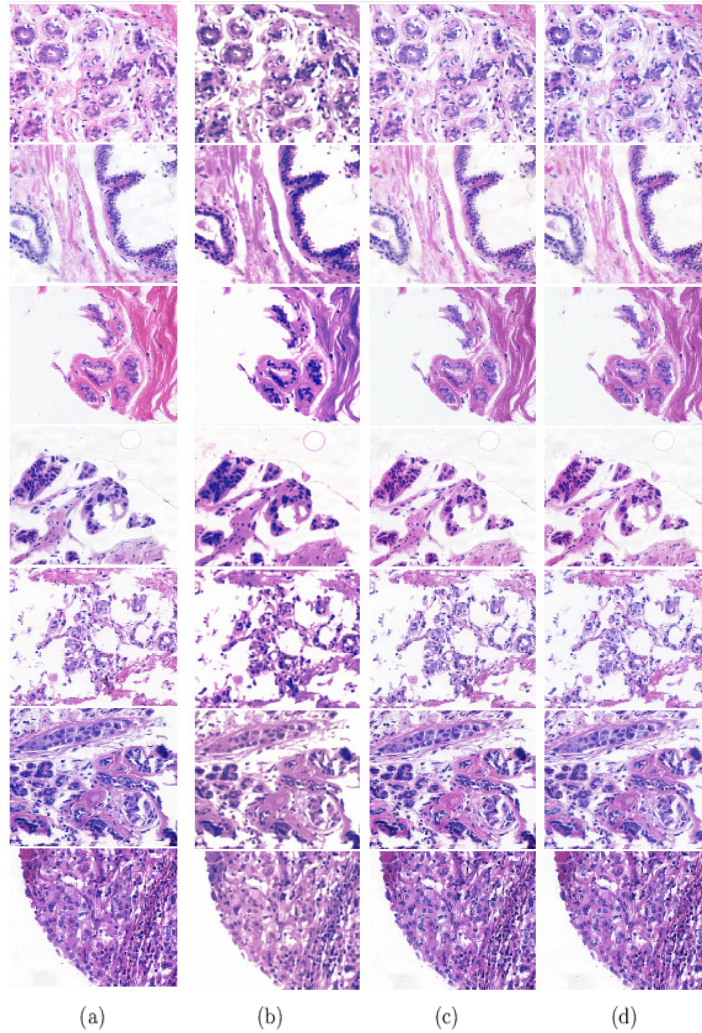
Figure 5.17: (a) Original images of UCSB data set; and color normalized images obtained using different state-of-the-art color normalization methods: (b) AST, (c) RFCC$_{vM}$ and (d) TredMiL.

image and the corresponding color normalized image is high, that is, SSIM value is close to 1, then it can be concluded that the structural information is preserved after color normalization.

From Table 5.5, it is evident that the proposed TredMiL model achieves highest mean and median values with respect to SSIM. The low standard deviation also highlights the fact that the TredMiL model produces consistently high values for all the images of UCSB data set. The TredMiL performs better than the existing methods due to the fact that the independence assumption between color appearance and stain density information helps the model to preserve the structural information. The color appearance information of the template image, rather than that of the source image, is fed into the decoder/generator during mapping, along with the stain density information of the source image. On the other hand, existing methods lose some amount of structural information during the afore-
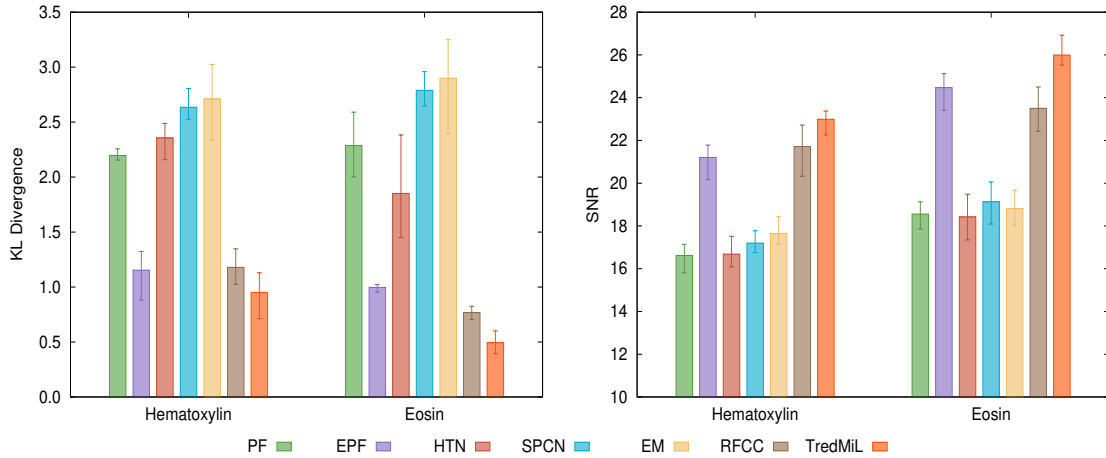
Figure 5.18: Performance analysis of different algorithms on CMU data using mean value of each index with error bar. A good stain separation method produces lower value of symmetric KL divergence and higher value of SNR. Here, RFCC refers to RFCC_{vM} method.

mentioned process as the color appearance and stain density information are assumed to be correlated.

5.4 Conclusion and Future Directions

In histological image analysis, the primary and foremost task is color normalization as it has an influence on the performance of the subsequent image analysis steps. The color variation present among images may lead to faulty diagnosis of diseases under consideration. In this respect, the major contributions of this chapter are listed as follows:

- development of a new deep generative model for color normalization of histology images, which incorporates truncated normal mixture as latent color appearance prior;

- formulating the objective functions, corresponding to generative module and reconstructive module, to capture disentangled color appearance and stain bound information, based on the assumption that latent color appearance code and latent stain density information are independent;

- modeling the discriminator to distinguish between real encoding and generated/fake encoding where elements of input triplets are sampled individually from different source distributions;

- applicability and generalizability of the trained model to histology image sets associated with same type of stains;

- demonstrating the efficacy of the proposed model, along with a comparison with state-of-the-art algorithms, on publicly available H&E stained histology image sets.

94
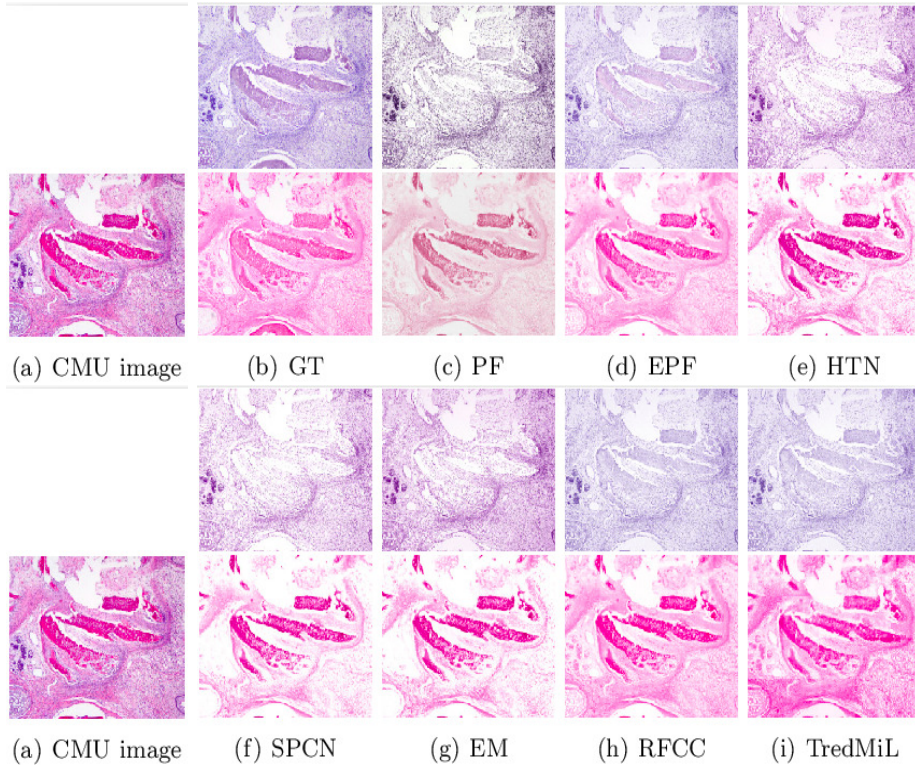
Figure 5.19: CMU image, ground-truth stain spectra, and estimated stain spectra obtained using different stain separation methods (top: H-stain; bottom: E-stain). In this figure, RFCC refers to $RFCC_{vM}$ method.

The proposed method works on image patches directly, not on any approximated histogram, which is not capable of capturing any spatial relation. The proposed TredMiL model deals with the overlapped appearance of histochemical reagents, by using truncated normal mixture model as the latent color appearance prior. From the quantitative as well as qualitative results, it can be seen that the proposed model outperforms state-of-the-art stain estimation and stain separation algorithms. The proposed method performs better than other existing color normalization methods, in terms of handling overlapping information, and ensuring within-image and between-image color constancy of ROI after color normalization. Also, the TredMiL model does not need any prior information except the number of stains involved in the staining routine. The trained model can be applied to any data set with H&E staining, which is one of the most commonly used staining methods in histopathological study.

To examine the applicability of the proposed TredMiL model on other types of staining, Hematoxylin-Diaminobenzidine (H&DAB) stained Warwick beta cell images [83] are considered in this chapter, and the corresponding results are reported in Fig. 5.20. The Warwick data set contains only 20 images, out of which patches from 12 images are used for training, 2 images are used in validation and image patches from 6 images are used for testing. From the qualitative results presented in Fig. 5.20, it can be observed that the model is capable of highlighting some biological components and reduces the variation

(a) patch 1   (b) patch 2   (c) patch 3   (d) patch 4
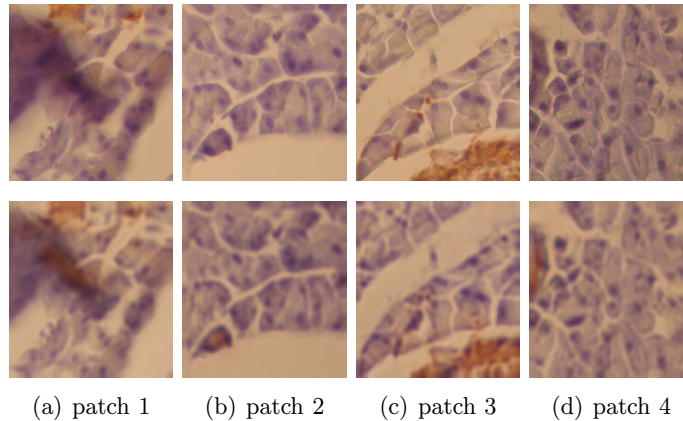
Figure 5.20: Performance analysis of TredMiL model on H&DAB stained Warwick beta cell images. Row 1 presents the image patches from Warwick data set, whereas Row 2 presents the corresponding color normalized image patches.

in the color appearance of the histological images, but the lack of substantial amount of samples hinders the proper training of the proposed TredMiL model. So, for other types of staining, the model should be trained with sufficient number of samples from at least one data set of that particular stain type. Also, the brown DAB in H&DAB staining is not a light absorber, it basically scatters light. So, a thorough understanding of immuno-histochemical stained histological slides is required before applying the model to H&DAB stained image data sets.

This chapter demonstrates that color normalization has a huge impact on subsequent nuclei segmentation task. Mitigating the disagreement in the appearance of stained tissue color improves quality of region-of-interest (ROI) segmentation from histological images. After addressing the color normalization problem, the segmentation of regions-of-interest from histological images is a very crucial task for computer-aided histological image analysis technique. To devise a nuclei segmentation algorithm, generalizable to segmentation problems with different types of target map representations, image-to-image translation paradigm can be considered as they are invariant to target domain image representations. But, image-to-image translation models fail miserably when the information content between two image domains are asymmetric in nature. In this regard, the next chapter introduces a new deep generative model for segmenting nuclei structures from histological images by handling the information-asymmetry between two image domains.

Chapter 6

Optimal Transport Driven Asymmetric Image-to-Image Translation for Nuclei Segmentation

6.1 Introduction

Histological tissue slides play a very crucial role in the diagnosis, prognosis, and treatment of cancer as they contain an abundance of phenotypic information [55]. With the advent of digital pathology, computer-aided analysis of histological tissue slides has become an important topic in the field of medical image analysis as it can ease pathologists' tasks and improve the accuracy of diagnosis. To help the pathologists in the microscopic analysis of tissue samples, slides are stained with several types of histochemical staining agents, such as hematoxylin and eosin (H&E), which eventually highlight different tissue components and nuclei morphology. The gradation of different types of cancer in patients is done by analyzing different shapes and spatial arrangements of nuclei within a tissue specimen [102]. Hence, segmentation of nuclei structures from histological images has an immense importance in digital pathology as it provides significant morphological information, which helps in therapeutic diagnosis and assessment of cancer. Nuclei segmentation can be defined as a process that focuses on the labeling of all pixels corresponding to each individual nucleus, and separating each nucleus from the background and other cellular components in a tissue slide image. Segmenting cell nuclei from histological tissue images is a challenging task as nuclei can exhibit different morphologies, color, texture or can be occluded partially by other nuclei or cellular components.

Segmentation and subsequently classification of nuclei are integral steps in histological image analysis. However, nuclei segmentation has gained a lot of attention in the last few years as accurate segmentation can facilitate downstream image analysis tasks. Some of the earlier works in this field have relied on energy-based methods, particularly watershed algorithm. In [92], the authors have extracted markers using fast radial symmetry transform and utilized a sequence of morphological operations to extract energy landscape. In [112], the authors have proposed a variant of the watershed algorithm that solely de-

pends on energy landscape, and uses a combination of active contour and shape prior to extract nuclear regions. Utilizing the geometry of nuclear structures, a method has been proposed in [62], which computes the concavity of nuclei clusters. A cell/nuclei segmentation approach, based on the combination of double thresholding method and morphological operations, has been proposed in [29]. However, most of these methods are sensitive to the choice of pre-defined parameters and assume the shape of the nuclei apriori, which is not a legitimate assumption as nuclei can exhibit diverse shapes across different organs. To tackle pathological image segmentation problem, some shallow learning based models have also been proposed in [61], [60], which perform convolution operation directly on the input histological image and try to extract pixel-wise labels through their abstract feature representations. The main issue with these approaches is that they do not consider spatial dependencies between pixels and hence capture inadequate information.

Recent years have witnessed a huge surge in the application of deep learning methods in many medical image analysis tasks. In [40], Ronneberger *et al.* have proposed U-Net model, a variant of fully convolutional network (FCN), which has been applied in numerous biomedical image segmentation tasks. The U-Net architecture incorporates the concept of skip connections in traditional encoder-decoder framework to capture low-level semantic information. The U-Net++, an extension of the U-Net architecture, has been introduced in [150], which enables feature propagation through densely interconnected skip connections. In [113], the authors have extended the U-Net architecture and proposed Micro-Net for proper detection and segmentation of nuclei of different sizes. However, U-Net based methods often fail to detect clustered nuclei and are extremely sensitive to pre-defined parameters associated with weighted loss function. A supervised two-stage object segmentation method, called Mask-RCNN, has been proposed in [66] for segmenting objects from natural as well as medical images. The model predicts bounding boxes corresponding to each individual nucleus, and then, the nuclei regions are segmented inside the predicted bounding boxes. In [114], the authors have proposed a model, named HoVer-Net, a unified FCN framework for simultaneous segmentation and classification of nuclei structures. To separate occluded or touching nuclei, vertical and horizontal distance maps have been utilized in this method. Most of these approaches need a large number of annotated images, which is sometimes practically infeasible due to the time, effort and expert knowledge needed to annotate the data. In [93], the authors have proposed a nuclei instance segmentation and classification approach, named Stardist. Though Stardist has attained much attention in recent times and performs well in case of convex shapes, it fails to detect heterogeneous nuclei shapes due to the fact that it assumes the shapes of the objects to be star-convex, which is not a valid assumption about nuclei structures belonging to different human organs.

In recent years, weakly supervised learning based nuclei segmentation methods have been introduced [53], [79], [148] that utilize weak point annotations for supervision. In [79], the authors have developed a modular deep network to decouple semantic and instance segmentation, and then fuse these two information to perform instance-aware nuclei segmentation. The main issue with this approach is that it requires both semantic and instance segmentation map corresponding to each training set histological image. In another approach [148], the authors have proposed a transformer based multiple instance learning method for generating prediction masks. However, weakly supervised learning based nuclei segmentation models fail to exploit image-level annotations and eventually miss some

of the nuclei instances. Recently, in [140], a weakly supervised nuclei segmentation method, called BoNuS has been proposed, which proposes a binary mining loss term to simultaneously learn nuclei interior and extract boundary information in order to extract nuclei structures from histological images. In [43], the authors have employed conditional GAN (cGAN) with spectral normalization and gradient penalty in CycleGAN [64] framework for nuclei segmentation of histological images. A pathology-constrained CycleGAN model has been proposed in [117], which can be utilized to solve the nuclei segmentation problem. However, these models are computationally very expensive as they incorporate one generator-discriminator pair corresponding to each individual domain.

Image-to-image (I2I) translation models aim to learn mappings between image domains while preserving certain high-level semantic information. One important characteristic of I2I models is that they are invariant to different types of target domain image representations, which enhances the applicability and generalizability of the models. To achieve unpaired training between two image domains, CycleGAN model was proposed in [64], which introduced cycle-consistency loss terms along with two generators: one for the forward mapping and other one for inverse mapping. The symmetric structure in CycleGAN model generalizes well and generates realistic images where the two image domains have similar information complexity. However, CycleGAN fails miserably when the information content within the two image domains are asymmetric in nature. As histological images contain information regarding cell nuclei along with other tissue-level details, they form information-rich image domain, compared to segmentation map space, which contains information regarding only nuclei structures. Hence, an image-to-image translation model needs to be developed to solve the nuclei segmentation problem by handling the information disparity among two image domains.

In this regard, a new deep generative model, termed as Ostrich (Optimal Transport Driven Asymmetric Image-to-Image Translation approach), is introduced in this chapter to segment nuclei structures from histological images. The model considers an embedding space for handling information-asymmetry between information-rich histological image space and information-poor segmentation map domain. The current study shows that the concept of invertible generators can automatically eliminate the need of any implicit cycle-consistency loss term. Integrating judiciously the concepts of optimal transport and measure theory, the model develops an invertible generator, which provides an efficient optimization framework with lower network complexity. The proposed model introduces a spatially-constrained squeeze operation within the framework of coupling layers to maintain intra-image spatial continuity. The proposed model provides a better trade-off between network complexity and model performance compared to other image-to-image translation models having more complex network architectures. The efficacy of the proposed model, along with a comparison with state-of-the-art nuclei segmentation approaches, has been demonstrated on publicly available histological image data sets.

The the rest of this chapter is organized as follows: Section 6.2 introduces a novel deep generative model, along with optimal transport and measure theory based theoretical analysis, to address the segmentation of nuclei structures from histological images. The performance of the proposed deep generative model is compared with several existing nuclei segmentation approaches on publicly available histological image data sets in Section 6.3. The concluding remarks are provided in Section 6.4.

99

6.2 Ostrich: Proposed Method

In this section, a deep generative model, based on the theory of optimal transport [30], is presented for nuclei segmentation from histological images. The proposed model is inspired by the cycle-free CycleGAN model introduced in [124].

6.2.1 Problem Statement

Given a histological image domain $\mathcal{X}$, containing a set of n number of images $\{x_i : i = 1, 2, \cdots, n\}$, the aim is to develop a model, which takes each histological image $x_i \in \mathcal{X} \subset \mathbb{R}^{H \times W \times 3}$ as input and outputs corresponding nuclei segmentation map y_i. The set of these output maps $\{y_i\}$ forms segmentation map domain $\mathcal{Y}$, that is, $y_i \in \mathcal{Y} \subset \mathbb{R}^{H \times W \times 3}$ (in case of instance segmentation) or $y_i \in \mathcal{Y} \subset \mathbb{R}^{H \times W}$ (in case of binary semantic segmentation). As described in the previous section, there is an observable information-imbalance between the two image domains: histological image space $\mathcal{X}$ and segmentation map domain $\mathcal{Y}$. Hence, the goal of the current study is to design a deep generative model that upon proper training will be able to generate segmentation annotation map corresponding to each input histological image by handling the information-asymmetry between two image domains.

6.2.2 Asymmetric Cycle-free Image-to-Image Translation

This subsection introduces an invertible image-to-image translation approach for solving the problem of nuclei segmentation of histological images. The following analysis establishes that two generators and one discriminator are sufficient to solve this problem.

To deal with the additional information while translating from information-rich image domain $\mathcal{X}$ to information-poor annotation image space $\mathcal{Y}$, an additional embedding space $\mathcal{Z}$ is considered in this study. Let, the histological image space $\mathcal{X}$ be equipped with probability measure μ, while the annotation image space $\mathcal{Y}$ and the embedding space $\mathcal{Z}$ be equipped with probability measures ν and η, respectively. The aim of the proposed model is to transport the joint distribution of ν and η of the annotation domain $\mathcal{Y}$ and embedding space $\mathcal{Z}$, respectively, that is, $(\nu \times \eta)$, to distribution μ of the image space $\mathcal{X}$ so that the joint distribution can mimic the distribution of the histological image space.

In the proposed model, there are eventually three generators F_ϕ, E_ω, and G_θ. The forward operator F_ϕ "pushes forward" measure μ in image space $\mathcal{X}$ to measure ν_ϕ in the annotation space $\mathcal{Y}$, that is, $F_\phi : \mathcal{X} \to \mathcal{Y}$. Similarly, the transportation from probability space $(\mathcal{X}, \mu)$ to $(\mathcal{Z}, \eta)$ is performed by the forward operator E_ω, which pushes forward measure μ in image space $\mathcal{X}$ to η_ω in the embedding space $\mathcal{Z}$, that is, $E_\omega : \mathcal{X} \to \mathcal{Z}$. On the other hand, the generator G_θ transports the product measure $(\nu \times \eta)$ in the product space $(\mathcal{Y} \times \mathcal{Z})$ to μ_θ in the target histological image space $\mathcal{X}$, that is, $G_\theta : (\mathcal{Y} \times \mathcal{Z}) \to \mathcal{X}$. Hence, the optimal transport problem in this study can be solved by minimizing the distances $d(\mu, \mu_\theta)$, $d(\nu, \nu_\phi)$, and $d(\eta, \eta_\omega)$, where $d(\alpha, \beta)$ denotes statistical distance between two probability measures α and β. So, to solve this simultaneous statistical distance minimization problem, the following optimal transport problem needs to be solved:

$$\inf_{\pi \in \Pi(\mu, \nu, \eta)} \int_{\mathcal{X} \times \mathcal{Y} \times \mathcal{Z}} \mathcal{C}(x, y, z; G_\theta, F_\phi, E_\omega) d\pi(x, y, z) \tag{6.1}$$

100

where $\Pi(\mu, \nu, \eta)$ denotes the set of all joint probability distributions with the marginals μ, ν, and η. In this study, the transportation cost is defined as follows:

$$C(x, y, z; G_\theta, F_\phi, E_\omega) = \|x - G_\theta(y, z)\| + \|F_\phi(x) - y\| + \|E_\omega(x) - z\|. \quad (6.2)$$

At the initial stage, the output segmentation maps, generated by the forward operator F_ϕ, are far from the real data distribution in $\mathcal{Y}$. So, to push the generation in the proper direction, an additional regularization term is introduced in the current study, which is defined as follows:

$$l_{ssim}(F_\phi) = \|1 - SSIM(F_\phi(x), VO(f_{rgb}^{hed}(x)[:,:,0]))\| \quad (6.3)$$

Here, $SSIM$ denotes a structural similarity measurement index, called structural similarity index measure [129], $f_{rgb}^{hed}(.)$ represents a function for converting a color image from RGB to Hematoxylin-Eosin-DAB (HED) colorspace. As nuclei regions in a H&E-stained image are generally highlighted by hematoxylin (H) stain, the H-channel of an image $x \in \mathcal{X}$ is considered through $f_{rgb}^{hed}(x)[:,:,0]$. $VO(.)$ denotes a transformation, which takes the H-channel image as input and outputs corresponding segmentation map by leveraging Voronoi labeling on thresholded H-channel image due to Otsu [99]. The aim is to maximize the structural similarity between the generated output and the transformed output through this regularization term. So, considering the regularization term, the overall transportation cost is defined as follows:

$$C(x, y, z; G_\theta, F_\phi, E_\omega) = \|x - G_\theta(y, z)\| + \|F_\phi(x) - y\| + \|E_\omega(x) - z\| + \lambda_1 l_{ssim}(F_\phi) \quad (6.4)$$

where λ_1 denotes the relative importance of the regularization term $l_{ssim}(F_\phi)$, defined in (6.3). The value of λ_1 decreases gradually during the training. The computation of the regularization term is trivial, but from the perspective of optimal transport, the computation of the first three terms requires corresponding dual formulation [30], [45].

The derivation of the dual formulation corresponding to the primal form presented in (6.1) is provided here:

$$\inf_{\pi \in \Pi(\mu, \nu, \eta)} \int_{\mathcal{X} \times \mathcal{Y} \times \mathcal{Z}} C(x, y, z; G_\theta, F_\phi, E_\omega) d\pi(x, y, z) \quad (6.5)$$

where the transportation cost $C(x, y, z; G_\theta, F_\phi, E_\omega)$ is defined in (6.4). So, (6.5) can be reformulated as follows:

$$\inf_{\pi \in \Pi(\mu, \nu, \eta)} \int_{\mathcal{X} \times \mathcal{Y} \times \mathcal{Z}} \hat{C}(x, y, z) d\pi(x, y, z) + \lambda_1 l_{ssim}(F_\phi) \quad (6.6)$$

where $l_{ssim}(F_\phi)$ is defined in (6.3) and

$$\hat{C}(x, y, z) = \|x - G_\theta(y, z)\| + \|F_\phi(x) - y\| + \|E_\omega(x) - z\|. \quad (6.7)$$

To compute the dual formulation of the above primal, let

$$K(G_\theta, F_\phi, E_\omega) = \min_{\pi \in \Pi(\mu, \nu, \eta)} \int_{\mathcal{X} \times \mathcal{Y} \times \mathcal{Z}} \hat{C}(x, y, z) d\pi(x, y, z). \tag{6.8}$$

Let π^* be the optimal joint measure, that is, $\pi^* = \arg\min_{\pi}\{\pi \in \Pi(\mu, \nu, \eta)\}$, then (6.8) can be rewritten as follows:

$$K(G_\theta, F_\phi, E_\omega) = \int_{\mathcal{X} \times \mathcal{Y} \times \mathcal{Z}} \hat{C}(x, y, z) d\pi^*(x, y, z). \tag{6.9}$$

Using (6.7), Kantorovich duality theorem [30] and c-transforms of Kantorovich potentials φ, ψ, and ξ, (6.9) can be rewritten as follows:

$$
\begin{aligned}
K(G_\theta, F_\phi, E_\omega) &= \int_{\mathcal{X} \times \mathcal{Y} \times \mathcal{Z}} \{\|x - G_\theta(y, z)\| + \|F_\phi(x) - y\| + \|E_\omega(x) - z\|\} d\pi^*(x, y, z) \\
&= \frac{1}{3}\{\max_\varphi[\int_{\mathcal{X}} \varphi(x) d\mu(x) + \int_{\mathcal{Y} \times \mathcal{Z}} \inf_x[\|x - G_\theta(y, z)\| + \|F_\phi(x) - y\| + \|E_\omega(x) - z\| - \varphi(x)] \\
&\qquad\qquad\qquad\qquad d(\nu \times \eta)(y, z)] \\
&+ \max_\psi[\int_{\mathcal{Y}} \psi(y) d\nu(y) + \int_{\mathcal{X}} \inf_y[\|x - G_\theta(y, z)\| + \|F_\phi(x) - y\| \|E_\omega(x) - z\| - \psi(y)] d\mu(x)] \\
&+ \max_\xi[\int_{\mathcal{Z}} \xi(z) d\eta(z) + \int_{\mathcal{X}} \inf_z[\|x - G_\theta(y, z)\| + \|F_\phi(x) - y\| + \|E_\omega(x) - z\| - \xi(z)] d\mu(x)]\}.
\end{aligned}
\tag{6.10}
$$

It can be noted from the optimization perspective that $\|E_\omega(x) - z\|$ and $\|F_\phi(x) - y\|$ are not dependant on $\mathcal{Y}$ and $\mathcal{Z}$ space, respectively.

Now, instead of finding $\inf_x$, $x = G_\theta(y, z) = G_{\theta y}(z) = G_\theta^z(y)$ is chosen. Again, instead of finding $\inf_y$ and $\inf_z$, if $y = F_\phi(x)$, and $z = E_\omega(x)$, respectively, are chosen, then an upper bound of $K(G_\theta, F_\phi, E_\omega)$ can be obtained from (6.10) as follows:

$$
\begin{aligned}
K(G_\theta, F_\phi, E_\omega) &\leqslant \frac{1}{3}\{\max_\varphi[\int_{\mathcal{X}} \varphi(x) d\mu(x) - \int_{\mathcal{Y} \times \mathcal{Z}} \varphi(G_\theta(y, z)) d(\nu \times \eta)(y, z)] \\
&+ \max_\psi[\int_{\mathcal{Y}} \psi(y) d\nu(y) - \int_{\mathcal{X}} \psi(F_\phi(x)) d\mu(x)] + \max_\xi[\int_{\mathcal{Z}} \xi(z) d\eta(z) - \int_{\mathcal{X}} \xi(E_\omega(x)) d\mu(x)] \\
&+ \int_{\mathcal{Y} \times \mathcal{Z}} [\|y - F_\phi(G_\theta(y, z))\| + \|z - E_\omega(G_\theta(y, z))\|] d(\nu \times \eta)(y, z)
\end{aligned}
$$

$$+ \int_{\mathcal{X}} [\|x - G_\theta(F_\phi(x), z)\| + \|x - G_\theta(y, E_\omega(x))\|] d\mu(x)\}$$

$$\Rightarrow K(G_\theta, F_\phi, E_\omega) \leqslant \frac{1}{3} \{l_{GAN}(G_\theta, F_\phi, E_\omega; \varphi, \psi, \xi) + l_{cycle}(G_\theta, F_\phi, E_\omega)\} \qquad (6.11)$$

where, $l_{GAN}(G_\theta, F_\phi, E_\omega; \varphi, \psi, \xi)$ and $l_{cycle}(G_\theta, F_\phi, E_\omega)$ can be defined as follows:

$$l_{GAN}(G_\theta, F_\phi, E_\omega; \varphi, \psi, \xi) = \max_{\varphi \in Lip_1(\mathcal{X})} \left[\int_{\mathcal{X}} \varphi(x) d\mu(x) - \int_{\mathcal{Y} \times \mathcal{Z}} \varphi(G_\theta(y, z)) d(\nu \times \eta)(y, z) \right]$$

$$\max_{\psi \in Lip_1(\mathcal{Y})} \left[\int_{\mathcal{Y}} \psi(y) d\nu(y) - \int_{\mathcal{X}} \psi(F_\phi(x)) d\mu(x) \right] + \max_{\xi \in Lip_1(\mathcal{Z})} \left[\int_{\mathcal{Z}} \xi(z) d\eta(z) - \int_{\mathcal{X}} \xi(E_\omega(x)) d\mu(x) \right]$$

$$(6.12)$$

and,

$$l_{cycle}(G_\theta, F_\phi, E_\omega) = \int_{\mathcal{X}} \|x - G_\theta{}^z(F_\phi(x))\| d\mu(x) + \int_{\mathcal{X}} \|x - G_{\theta y}(E_\omega(x))\| d\mu(x)$$

$$+ \int_{\mathcal{Y}} \|y - F_\phi(G_\theta{}^z(y))\| d\nu(y) + \int_{\mathcal{Z}} \|z - E_\omega(G_{\theta y}(z))\| d\eta(z) \qquad (6.13)$$

In (6.12), $Lip_m(D)$ represents the space of all m-Lipschitz functions defined over domain D. G_θ is a $(\nu \times \eta)$-measurable function defined on the product space $(\mathcal{Y} \times \mathcal{Z})$. So, the sections of G_θ, that is, $G_{\theta y}$ and $G_\theta{}^z$, are η-measurable function defined on $\mathcal{Z}$ space and ν-measurable function defined on domain $\mathcal{Y}$, respectively [33].

Again, using 1-Lipschitz continuity of Kantorovich potentials φ, ψ, and ξ, the following relation can be obtained:

$$\varphi(x) - \varphi(G_\theta(y, z)) \leqslant \|x - G_\theta(y, z)\| \leqslant \|x - G_\theta(y, z)\| + \|F_\phi(x) - y\| + \|E_\omega(x) - z\|$$
$$\Rightarrow -\varphi(G_\theta(y, z)) \leqslant \|x - G_\theta(y, z)\| - \varphi(x)$$
$$\leqslant \|x - G_\theta(y, z)\| + \|F_\phi(x) - y\| + \|E_\omega(x) - z\| - \varphi(x). \qquad (6.14)$$

Similarly, the following relations can be obtained:

$$-\psi(F_\phi(x)) \leqslant \|F_\phi(x) - y\| - \psi(y) \leqslant \|x - G_\theta(y, z)\| + \|F_\phi(x) - y\| + \|E_\omega(x) - z\| - \psi(y)$$
$$(6.15)$$

and

$$-\xi(E_\omega(x)) \leqslant \|E_\omega(x) - z\| - \xi(z) \leqslant \|x - G_\theta(y, z)\| + \|F_\phi(x) - y\| + \|E_\omega(x) - z\| - \xi(z)$$
$$(6.16)$$

The inequalities (6.14)-(6.16) lead to the following lower bound:

$$K(G_\theta, F_\phi, E_\omega) \geqslant \frac{1}{3} l_{GAN}(G_\theta, F_\phi, E_\omega; \varphi, \psi, \xi). \tag{6.17}$$

It is evident from (6.11) and (6.17) that if $l_{cycle}(G_\theta, F_\phi, E_\omega)$ vanishes, then the following relation can be obtained:

$$K(G_\theta, F_\phi, E_\omega) = \frac{1}{3} l_{GAN}(G_\theta, F_\phi, E_\omega; \varphi, \psi, \xi). \tag{6.18}$$

The dual formulation corresponding to the optimal transport problem presented in (6.1), with the transportation cost $C(x, y, z; G_\theta, F_\phi, E_\omega)$, defined in (6.4), can be represented as follows:

$$\min_{\theta, \phi, \omega} \max_{\varphi, \psi, \xi} l(G_\theta, F_\phi, E_\omega; \varphi, \psi, \xi) \tag{6.19}$$

where φ, ψ and ξ represent three discriminators corresponding to three domains $\mathcal{X}$, $\mathcal{Y}$ and $\mathcal{Z}$, respectively. Here, the loss $l(G_\theta, F_\phi, E_\omega; \varphi, \psi, \xi)$ is defined as follows:

$$l(G_\theta, F_\phi, E_\omega; \varphi, \psi, \xi) = l_{GAN}(G_\theta, F_\phi, E_\omega; \varphi, \psi, \xi) + \lambda_2 l_{cycle}(G_\theta, F_\phi, E_\omega) + \lambda_1 l_{ssim}(F_\phi) \tag{6.20}$$

The parameter λ_2 in (6.20) denotes the relative importance of the cycle-consistency loss term $l_{cycle}(G_\theta, F_\phi, E_\omega)$, which is defined in (6.13).

Theorem 6.1. *If the sections of generator G_θ, that is, $G_{\theta y}$ and $G_\theta{}^z$, are considered to be the inverse operators of forward operators E_ω and F_ϕ, respectively, that is, $G_{\theta y} = E_\omega{}^{-1}$ and $G_\theta{}^z = F_\phi{}^{-1}$ and $G_\theta{}^z$ is a p-Lipschitz functions in domain $\mathcal{Y}$, then the problem in (6.19) with $p = 1$, can be represented by the following equivalent problem:*

$$\min_{\phi, \omega} \max_{\psi, \xi} l(F_\phi, E_\omega; \psi, \xi) \tag{6.21}$$

where $\quad l(F_\phi, E_\omega; \psi, \xi) = 2 l_{GAN}(F_\phi; \psi) + l_{GAN}(E_\omega; \xi) + \lambda_1 l_{ssim}(F_\phi)$ $\quad$ (6.22)

and

$$l_{GAN}(F_\phi; \psi) = \max_{\psi \in Lip_1(\mathcal{Y})} \int_\mathcal{Y} \psi(y) d\nu(y) - \int_\mathcal{X} \psi(F_\phi(x)) d\mu(x), \tag{6.23}$$

$$l_{GAN}(E_\omega; \xi) = \max_{\xi \in Lip_1(\mathcal{Z})} \int_\mathcal{Z} \xi(z) d\eta(z) - \int_\mathcal{X} \xi(E_\omega(x)) d\mu(x). \tag{6.24}$$

Proof. The invertibility conditions $G_{\theta y} = E_\omega{}^{-1}$ and $G_\theta{}^z = F_\phi{}^{-1}$ ensure that $G_{\theta y}(E_\omega(x)) = x$, $G_\theta{}^z(F_\phi(x)) = x$, $F_\phi(G_\theta{}^z(y)) = y$ and $E_\omega(G_{\theta y}(z)) = z$. So, it is evident that if the above invertibility conditions hold, then the cycle-consistency loss term $l_{cycle}(G_\theta, F_\phi, E_\omega)$ vanishes.

The first sub-expression from (6.12) is considered as follows:

$$\max_{\varphi \in Lip_1(\mathcal{X})} \left[\int_{\mathcal{X}} \varphi(x)d\mu(x) - \int_{\mathcal{Y} \times \mathcal{Z}} \varphi(G_\theta(y, z))d(\nu \times \eta)(y, z) \right]. \tag{6.25}$$

Let, $\varphi^* \in Lip_1(\mathcal{X})$ be the optimum 1-Lipschitz function in domain $\mathcal{X}$ for which the above expression attains maximum value. So, (6.25) can be rewritten as follows:

$$\max_{\varphi \in Lip_1(\mathcal{X})} \left[\int_{\mathcal{X}} \varphi(x)d\mu(x) - \int_{\mathcal{Y} \times \mathcal{Z}} \varphi(G_\theta(y, z))d(\nu \times \eta)(y, z) \right]$$
$$= \int_{\mathcal{X}} \varphi^*(x)d\mu(x) - \int_{\mathcal{Y} \times \mathcal{Z}} \varphi^*(G_\theta(y, z))d(\nu \times \eta)(y, z). \tag{6.26}$$

Let, the two terms in (6.26) be denoted by J_1 and J_2, respectively, that is,

$$J_1 = \int_{\mathcal{X}} \varphi^*(x)d\mu(x) \tag{6.27}$$

and

$$J_2 = - \int_{\mathcal{Y} \times \mathcal{Z}} \varphi^*(G_\theta(y, z))d(\nu \times \eta)(y, z). \tag{6.28}$$

As $G_\theta{}^z$ is designed to be the inverse of forward operator F_ϕ, the following formulation can be derived from (6.27):

$$J_1 = \int_{\mathcal{X}} \varphi^*(G_\theta{}^z(F_\phi(x)))d\mu(x) = \max_{\psi_\theta \in \Phi_1} \int_{\mathcal{X}} \psi_\theta(F_\phi(x))d\mu(x) \tag{6.29}$$

where the set Φ_1 is defined as follows:

$$\Phi_1 = \{\hat{\psi} | \hat{\psi} = \varphi \circ G_\theta{}^z, \varphi \in Lip_1(\mathcal{X}), G_\theta{}^z \in Lip_1(Y)\}. \tag{6.30}$$

Here, '$\circ$' denotes the composition of functions.

Now, (6.28) can be rewritten as follows:

$$J_2 = \max_{\mathcal{H}_\theta \in \Phi_2} \left[- \int_{\mathcal{Y} \times \mathcal{Z}} \mathcal{H}_\theta(y, z)d(\nu \times \eta)(y, z) \right]. \tag{6.31}$$

Here, the set Φ_2 is defined as follows:

$$\Phi_2 = \{\mathcal{H}|\mathcal{H} = \varphi \circ G_\theta, \varphi \in Lip_1(\mathcal{X})\}. \tag{6.32}$$

Using Fubini's theorem on product measures [33], (6.31) can be formulated as follows:

$$J_2 = \max_{\mathcal{H}_\theta \in \Phi_2} \left[-\int_{\mathcal{Y}} \left\{ \int_{\mathcal{Z}} \mathcal{H}_\theta(y, z)d\eta(z) \right\} d\nu(y) \right]. \tag{6.33}$$

The expression $\int_{\mathcal{Z}} \mathcal{H}_\theta(y, z)d\eta(z)$ in (6.33) is a ν-measurable function and is defined on space $\mathcal{Y}$. Let, the expression be denoted by $\mathcal{M}_\theta(y)$. For any two points $y_1, y_2 \in \mathcal{Y}$,

$$\|\mathcal{M}_\theta(y_1) - \mathcal{M}_\theta(y_2)\| = \|\int_{\mathcal{Z}} \mathcal{H}_\theta(y_1, z)d\eta(z) - \int_{\mathcal{Z}} \mathcal{H}_\theta(y_2, z)d\eta(z)\|$$

$$\leqslant \int_{\mathcal{Z}} \|\mathcal{H}_\theta(y_1, z) - \mathcal{H}_\theta(y_2, z)\|d\eta(z) \leqslant \int_{\mathcal{Z}} \|G_\theta(y_1, z) - G_\theta(y_2, z)\|d\eta(z)$$

$$[\text{as } \mathcal{H}_\theta = \varphi \circ G_\theta \text{ and } \varphi \in Lip_1(\mathcal{X})]$$

$$= \int_{\mathcal{Z}} \|G_\theta{}^z(y_1) - G_\theta{}^z(y_2)\|d\eta(z) \leqslant \|y_1 - y_2\| \int_{\mathcal{Z}} d\eta(z) = \|y_1 - y_2\| \quad [\text{as } G_\theta{}^z \in Lip_1(\mathcal{Y})].$$

$$\tag{6.34}$$

Thus, (6.33) can be reformulated as follows:

$$J_2 = \max_{\mathcal{M}_\theta \in \Phi_3} \left[-\int_{\mathcal{Y}} \mathcal{M}_\theta(y)d\nu(y) \right] \tag{6.35}$$

where the set Φ_3 can be defined as follows:

$$\Phi_3 = \{\mathcal{M}|\mathcal{M} = \int_{\mathcal{Z}} \mathcal{H}(y, z)d\eta(z), \mathcal{H} \in \Phi_2\}. \tag{6.36}$$

Hence, the set Φ_3 constructs a subset of the set of all 1-Lipschitz functions on domain $\mathcal{Y}$. Again, for any element $\hat{\psi}_1 \in \Phi_1$, $\hat{\psi}_1$ is of the form $\hat{\psi}_1 = \varphi_1 \circ g$, where $\varphi_1 \in Lip_1(\mathcal{X})$ and $g \in Lip_1(\mathcal{Y})$. So, g can potentially be any element from the set Φ_3 and $\hat{\psi}_1$ can map the element to $Lip_1(\mathcal{Y})$ within or beyond Φ_3. So, the above argument ensures that Φ_3 is contained in Φ_1, that is, $\Phi_3 \subseteq \Phi_1$. Hence, by combining (6.29) and (6.35), the following

relation can be deduced:

$$J_1 + J_2 \leqslant \max_{\psi' \in \Phi_1} \left[\int_{\mathcal{X}} \psi'(F_\phi(x)) d\mu(x) - \int_{\mathcal{Y}} \psi'(y) d\nu(y) \right]$$

$$= \max_{\tilde{\psi} \in \Phi_1} \left[\int_{\mathcal{Y}} \tilde{\psi}(y) d\nu(y) - \int_{\mathcal{X}} \tilde{\psi}(F_\phi(x)) d\mu(x) \right]$$

[equality comes from the fact: $\tilde{\psi} = -\psi'$ and thus $\tilde{\psi} \in \Phi_1$]. (6.37)

Now, by extending the function space from set Φ_1 to $Lip_1(\mathcal{Y})$, the following relation can be obtained from (6.25) and (6.37):

$$\max_{\varphi \in Lip_1(\mathcal{X})} \left[\int_{\mathcal{X}} \varphi(x) d\mu(x) - \int_{\mathcal{Y} \times \mathcal{Z}} \varphi(G_\theta(y,z)) d(\nu \times \eta)(y,z) \right]$$

$$\leqslant \max_{\psi \in Lip_1(\mathcal{Y})} \left[\int_{\mathcal{Y}} \psi(y) d\nu(y) - \int_{\mathcal{X}} \psi(F_\phi(x)) d\mu(x) \right]. \tag{6.38}$$

Next, it must be shown that the upper bound in (6.38) is tight. Let, ψ^* be the maximizer for (6.38). To show the tightness of the upper bound, it is important to show that there exists $\varphi \in Lip_1(\mathcal{X})$ such that, for $z \in \mathcal{Z}$:

$$\psi^*(y) = \varphi(G_\theta{}^z(y)) = \varphi(G_\theta(y,z)), \quad \forall y \in \mathcal{Y}.$$

Due to the invertibility condition $G_\theta{}^z = F_\phi{}^{-1}$, an element $x \in \mathcal{X}$ can always be found such that $x = F_\phi{}^{-1}(y) = G_\theta{}^z(y)$ for all $y \in \mathcal{Y}$. So,

$$\varphi(x) = \varphi(F_\phi{}^{-1}(F_\phi(x))) = \varphi(G_\theta{}^z(F_\phi(x))) = \psi^*(F_\phi(x)),$$

which achieves the above upper bound. Hence, the following interesting relation can be obtained from (6.38):

$$\max_{\varphi \in Lip_1(\mathcal{X})} \left[\int_{\mathcal{X}} \varphi(x) d\mu(x) - \int_{\mathcal{Y} \times \mathcal{Z}} \varphi(G_\theta(y,z)) d(\nu \times \eta)(y,z) \right]$$

$$= \max_{\psi \in Lip_1(\mathcal{Y})} \left[\int_{\mathcal{Y}} \psi(y) d\nu(y) - \int_{\mathcal{X}} \psi(F_\phi(x)) d\mu(x) \right]. \tag{6.39}$$

Substituting (6.39) in (6.12), following relation can be deduced:

$$l_{GAN}(G_\theta, F_\phi, E_\omega; \varphi, \psi, \xi) = \frac{1}{3} \{ 2l_{GAN}(F_\phi; \psi) + l_{GAN}(E_\omega; \xi) \} \tag{6.40}$$

which proves the claim of *Theorem 6.1*.

The above analysis eliminates the need of additional discriminator φ and generator G_θ if G_{θ_y} and $G_\theta{}^z$ are considered to be the inverse operators of forward operators E_ω and F_ϕ, respectively. From the optimization perspective, if the normalizing constants in (6.40) are not taken into consideration, then the model requires two generators F_ϕ and E_ω, and two discriminators ψ and ξ. As discriminators ψ and ξ discriminate between real and fake/generated $\mathcal{Y}$ domain data, and real and generated $\mathcal{Z}$ space data, respectively, instead of two discriminators, one discriminator ψ_ξ is sufficient, which takes inputs in the form of pairs (y, z). The discriminator $\psi_\xi(y, z)$ discriminates between the real pair (y, z), where $y \sim \mathcal{Y}$ and $z \sim \mathcal{Z}$, and the generated/fake pair (y, z), where $y \sim P_{F_\phi}(y \mid x)$ and $z \sim P_{E_\omega}(z \mid x)$ correspond to the outputs generated through F_ϕ and E_ω, respectively. If needed, a histological image x can be generated through the invertible generator $G_\theta{}^z$ for a given embedding information $z \in \mathcal{Z}$, corresponding to a segmentation map $y \in \mathcal{Y}$.
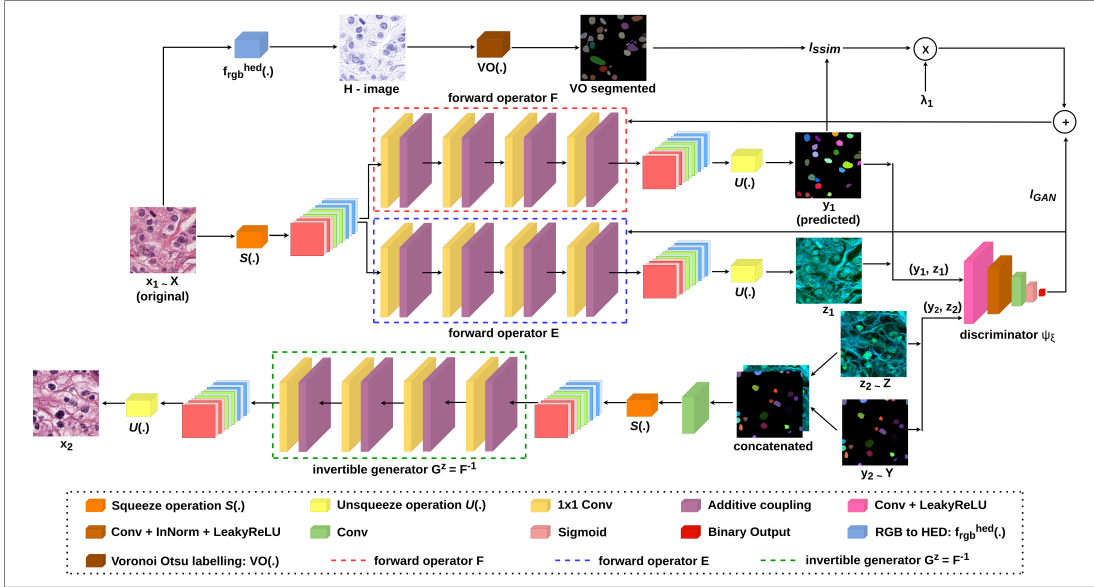
Figure 6.1: Block diagram of the proposed deep generative model for nuclei segmentation of histological images. Each network component is marked with distinct color representation. In the diagram, 'Conv' means convolutional layer, '1 × 1 Conv' refers to invertible 1 × 1 convolution approach proposed in [73]. Given $z_2 \in \mathcal{Z}$, z_2 is concatenated with y_2, and the concatenated image is fed to a convolution layer to produce a 3 channel output. 'InNorm' represents instance normalization, 'LeakyReLU' denotes leaky rectified linear unit. Here, forward operators F, E and invertible generator G^z refer to generators F_ϕ, E_ω and $G_\theta{}^z$, respectively.

6.2.3 Model Structure

Fig. 6.1 presents the block diagram of the proposed deep generative model for nuclei segmentation from histological images. The forward operator F_ϕ generates segmentation map corresponding to each input histological image and generator E_ω embeds additional
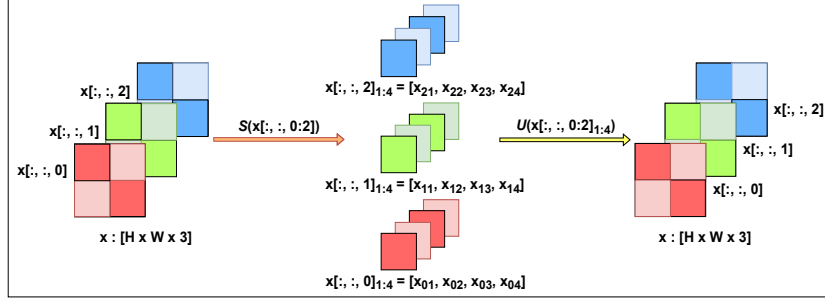
Figure 6.2: Block diagram of the squeeze-unsqueeze operation. Here, $x[:, :, 0 : 2]$ denotes 3 channel input image. $x[:, :, i]_{1:4}$ represents 4 sub-patches of the i-th channel of input image $x \in \mathcal{X}$.

information while translating from information-rich image space $\mathcal{X}$ to information-poor segmentation map domain $\mathcal{Y}$. The discriminator ψ_ξ discriminates between real and generated/fake pairs of segmentation map and embedding information. Each training image patch $x \in \mathcal{X}$ is fed into both forward operators $F_\phi(x; \Theta_\phi)$ and $E_\omega(x; \Theta_\omega)$, which output corresponding segmentation map y and embedding information z, respectively. Without loss of generality, $\mathcal{Z}$ is assumed to follow standard normal distribution. Here, Θ_ϕ, Θ_ω and Θ_{ψ_ξ} denote the network parameters associated with the networks F_ϕ, E_ω and ψ_ξ, respectively. As $G_\theta{}^z$ is designed to be the inverse of the forward operator F_ϕ, $G_\theta{}^z$ uses the same parameter set Θ_ϕ as that of F_ϕ.

The overall objective of the proposed model is to generate segmentation map corresponding to each input histological image by handling the information-asymmetry between the two domains. The following min-max objective function J_{adv} is to be optimized to achieve the aforementioned objective:

$$J_{adv} = J(\psi_\xi) + J(F_\phi, E_\omega) \tag{6.41}$$

where $J(\psi_\xi)$ is the maximization objective, attributed by the discriminator ψ_ξ, and is given by

$$J(\psi_\xi) = \max_{\psi_\xi} \{ l_{GAN}(F_\phi, E_\omega; \psi_\xi) \} \tag{6.42}$$

and $J(F_\phi, E_\omega)$ is the minimization objective, attributed by the two generators F_ϕ and E_ω, which is given by

$$J(F_\phi, E_\omega) = \min_{\phi, \omega} \{ l_{GAN}(F_\phi, E_\omega; \psi_\xi) + \lambda_1 l_{ssim}(F_\phi) \}$$
$$= \min_\phi \{ l_{GAN}(F_\phi, E_\omega; \psi_\xi) + \lambda_1 l_{ssim}(F_\phi) \} + \min_\omega \{ l_{GAN}(F_\phi, E_\omega; \psi_\xi) \} \tag{6.43}$$

where $l_{GAN}(F_\phi, E_\omega; \psi_\xi)$ can be defined as follows:

$$l_{GAN}(F_\phi, E_\omega; \psi_\xi) = \max_{\psi_\xi}[\int_{\mathcal{Y} \times \mathcal{Z}} \psi_\xi(y, z)d(\nu \times \eta)(y, z) - \int_{\mathcal{X}} \psi_\xi(F_\phi(x), E_\omega(x))d\mu(x)]. \quad (6.44)$$

The main building block of the invertible generator is implemented using the concepts of invertible 1×1 convolution, proposed in [73] and general coupling layer, introduced in [58]. Algorithm 6.1 summarizes the training procedure for the proposed Ostrich model. To maintain the spatial continuity within the image patches, a modified squeeze-unsqueeze operation is introduced in this study. For an input image $x \in \mathcal{X}$, the squeeze operation can be mathematically represented as follows:

$$[x[:,:,0]_{1:4}, x[:,:,1]_{1:4}, x[:,:,2]_{1:4}] = S(x[:,:,0:2]) = S(x)$$

whereas the unsqueeze operation can be defined as follows:

$$x = U([x[:,:,0:2]_{1:4}]).$$

The squeeze operation $S(.)$ and unsqueeze operation $U(.)$ are demonstrated schematically through Fig. 6.2.

6.3   Performance Analysis

This section analyzes the performance of the proposed deep generative model, termed as Ostrich.

6.3.1   Experimental Setup

A brief description of the existing algorithms compared, data sets and quantitative indices used, and details regarding model training is provided next.

6.3.1.1   Algorithms Compared

The performance of the proposed Ostrich model in nuclei segmentation is analyzed extensively and compared with that of several standard object segmentation models, such as, U-Net [40], Mask-R-CNN [66], U-Net++ [150]; and state-of-the-art nuclei segmentation methods, namely, HoVer-Net [114], multi-organ nuclei segmentation (MoNS) [43], Stardist [93], Swin-MIL [148] and BoNuS [140].

6.3.1.2   Data Sets Used

The performance of different variants of the proposed model and several state-of-the-art algorithms in nuclei segmentation is analyzed using the following two publicly available data sets.

Algorithm 6.1 Mini-batch stochastic gradient descent training of proposed Ostrich model.

Input: A set of training images sampled from real data distribution $P_X(x)$, segmentation map images sampled from real data distribution $P_Y(y)$, number of training epochs (τ).

Output: Trained parameter set $\{\Theta_{\psi_\xi}, \Theta_{F_\phi}, \Theta_{E_\omega}\}$, corresponding to three deep neural networks ψ_ξ, F_ϕ and E_ω, respectively.

1: **for** each epoch $t = 0$ to $\tau - 1$ **do**

- Sample mini-batch of M images $\{x_m\}_{m=1}^M$ from real data distribution $P_X(x)$.

- Generate M nuclei segmentation maps $\{y_m\}_{m=1}^M$ by using $F_\phi(x_m; \Theta_\phi)$.

- Generate M embedding information maps $\{z_m\}_{m=1}^M$ by using $E_\omega(x_m; \Theta_\omega)$.

- Sample mini-batch of M nuclei segmentation maps $\{\acute{y}_m\}_{m=1}^M$ from real segmentation map image distribution $P_Y(y)$.

- Sample mini-batch of M embedding information maps $\{\acute{z}_m\}_{m=1}^M$ from embedding information prior $P_Z(z)$.

- Update discriminator ψ_ξ by ascending in the direction of stochastic gradient:

$$
\nabla_{\Theta_{\psi_\xi}} \left(\frac{1}{M} \sum_{m=1}^M \left[\psi_\xi(\acute{y}_m, \acute{z}_m; \Theta_{\psi_\xi}) - \psi_\xi(y_m, z_m; \Theta_{\psi_\xi}) \right] \right)
$$

- Update segmentation map generator F_ϕ by descending in the direction of stochastic gradient:

$$
\nabla_{\Theta_{F_\phi}} \left(\frac{1}{M} \sum_{m=1}^M \left[-\psi_\xi(y_m, z_m; \Theta_{\psi_\xi}) + \lambda_1 \| 1 - SSIM(y_m, VO(f_{rgb}^{hed}(x_m)[:, :, 0])) \| \right] \right)
$$

- Update embedding map generator E_ω by descending in the direction of stochastic gradient:

$$
\nabla_{\Theta_{E_\omega}} \left(\frac{1}{M} \sum_{m=1}^M \left[-\psi_\xi(y_m, z_m; \Theta_{\psi_\xi}) \right] \right)
$$

2: **end for**

3: Stop.

- TCGA Data [97]: This multi-organ image data set contains 30 images of resolution 1000×1000, that were cropped from nuclei-dense regions of whole slide images downloaded from The Cancer Genome Atlas (TCGA). This H&E stained data set contains both instance segmentation and binary semantic segmentation maps corresponding to each histological image.

- CoNIC Data [115]: This H&E data set was published as a challenge data set: Colon Nuclei Identification and Counting (CoNIC) Challenge, 2022. This colon cancer tissue image data set consists of 4,981 image patches having resolution of 256×256. This data set contains instance segmentation maps corresponding to each histological image.

Four standard segmentation evaluation metrics, namely, Dice coefficient, Jaccard index, precision and recall, are used in this study to analyze the performance of different algorithms.

6.3.1.3 Training Details

The TCGA data set is randomly split into training, validation and test set with a ratio of $5 : 1 : 4$. The proposed Ostrich model is trained with $2,535$ overlapping image patches of size 256×256 corresponding to 15 training set images of TCGA data set, while the validation set contains 507 overlapping patches and the test set comprises of 192 non-overlapping image patches. The CoNIC data set consists of $4,981$ image patches, among which $2,930$ image patches have been used for training, 586 patches for validation and $1,465$ images have been used for testing the performance of the proposed model. The incorporation of validation set for both the data sets prevents the model from over-fitting and ensures early stopping based on validation loss.

6.3.2 Performance Analysis on TCGA Data

The effectiveness of the proposed Ostrich model is validated through the following analyses.

6.3.2.1 Ablation Study

The objective function of the proposed Ostrich model is comprised of two constituent terms: a GAN objective term and a regularization term. Let, the regularization term $l_{ssim}(F_\phi)$ be denoted by R. The importance of the regularization term R is analyzed by an ablation study, where R is omitted from the objective function by setting $\lambda_1 = 0$. The performance of the proposed Ostrich model in nuclei segmentation is then compared with that of the model in the absence of R, and the corresponding results are presented in Table 6.1. When λ_1 is set to 0, the model is referred to as "Ostrich $\backslash R$". From the results reported in Table 6.1, it is evident that, with respect to all the segmentation evaluation indices used in this study, the proposed model performs better than the "Ostrich $\backslash R$" model. The qualitative comparison presented in Fig. 6.3 ensures that, with the initial perturbation through regularization term R, the proposed model is able to capture intrinsic details of underlying data distribution. The statistical significance of the proposed Ostrich model is analyzed with respect to p-values computed through paired-t (one-tailed) and

112

Wilcoxon signed-rank (one-tailed) tests. From the p-values reported in Table 6.2 and 6.3, it can be observed that, with respect to both paired-t and Wilcoxon signed-rank tests, the proposed model performs significantly better than the "Ostrich $\backslash R$" model, considering 95% confidence level.

Table 6.1: Quantitative performance analysis in nuclei segmentation on TCGA Data set: State-of-the-art models vs Ostrich.

Methods	Dice	Jaccard	Precision	Recall
Ostrich	**0.788035**	**0.641520**	**0.813523**	0.764096
Ostrich $\backslash R$	0.761814	0.606573	0.791683	0.734116
CycleGAN	0.754221	0.598782	0.764761	0.743967
OT-CycleGAN	0.766668	0.614903	0.804106	0.732561
Asym-CycleGAN	0.783812	0.638815	0.806971	0.761945
Checkerboard	0.772020	0.629522	0.785028	0.759437
U-Net	0.645650	0.446128	0.624558	0.668216
Mask-R-CNN	0.747086	0.585659	0.799975	0.700757
U-Net++	0.773155	0.620294	0.796708	0.750955
HoVer-Net	0.744516	0.586780	0.811804	0.687529
MoNS	0.755509	0.614706	0.748012	0.763157
Stardist	0.743228	0.585713	0.811579	0.685495
Swin-MIL	0.749044	0.593266	0.801989	0.702656
BoNuS	0.784740	0.639496	0.806351	**0.764257**

| (a) | (b) | (c) | (d) |

Figure 6.3: Qualitative performance analysis in nuclei instance segmentation: (a) Original image, (b) Ground-truth; Segmentation performance of (c) Ostrich $\backslash R$ and (d) Ostrich. Top row presents marked-up images from TCGA data, and bottom row presents the zoomed-in region from the marked-up patches.

6.3.2.2 Comparison with Baseline Models

The performance of the proposed Ostrich model is compared with that of two existing image-to-image translation models, namely, CyclegGAN [64] and cycle-free Cycle-GAN [124]. To establish the significance of the proposed optimal transport and measure theory based analysis, the existing CycleGAN model is also extended by considering an embedding space for capturing additional information while translating from information-rich

Table 6.2: Statistical significance analysis of different algorithms with respect to Paired-t (one-tailed) test on TCGA Data set

Methods	Dice	Jaccard	Precision	Recall
Ostrich $\backslash R$	9.88E-30	3.41E-30	1.44E-09	1.30E-10
CycleGAN	4.88E-32	3.39E-34	3.69E-32	8.40E-10
OT-CycleGAN	2.90E-33	2.26E-33	1.04E-05	6.73E-21
Asym-CycleGAN	*6.49E-02*	*5.42E-02*	5.43E-05	*8.31E-02*
Checkerboard	4.66E-13	5.70E-14	1.83E-61	1.10E-02
U-Net	6.67E-30	4.18E-36	1.25E-76	6.68E-06
Mask-R-CNN	1.17E-37	4.55E-39	3.77E-05	1.28E-25
U-Net++	1.08E-17	3.86E-18	5.03E-09	7.46E-03
HoVer-Net	1.77E-41	7.41E-45	*2.37E-01*	7.65E-54
MoNS	1.10E-14	1.84E-15	3.34E-48	*1.21E-01*
Stardist	1.46E-24	8.26E-26	*3.09E-01*	9.93E-27
Swin-MIL	1.95E-37	1.64E-41	1.74E-03	3.25E-49
BoNuS	*1.42E-01*	*1.05E-01*	9.20E-07	*9.43E-01*

Figure 6.4: Performance analysis in nuclei segmentation: (a) Original histological image, (b) Ground-truth segmentation map; segmentation performance- (c) CycleGAN , (d) OT-CycleGAN, (e) Asym-CycleGAN and (f) Ostrich. Top row presents marked-up images from TCGA data, and bottom row presents the zoomed-in region from the marked-up patches.

histological image space to information-poor segmentation map domain. This asymmetric CycleGAN model is comprised of three generators and three discriminators and is referred to as Asym-CycleGAN in this section. For simplicity, the optimal transport driven cycle-free CycleGAN model is referred to as OT-CycleGAN. From the results reported in Table 6.1, it can be observed that the proposed Ostrich model outperforms baseline image-to-image translation models with respect to all the segmentation evaluation indices, namely, Dice, Jaccard, precision and recall. From the p-values reported in Table 6.2 and 6.3, it can be noted that the Ostrich model performs significantly better than CycleGAN and OT-CycleGAN, and better than Asym-CycleGAN model, but not significantly. It is also evident from Table 6.4 that the proposed Ostrich model uses lesser number of parameters compared to CycleGAN and Asym-CycleGAN models. Though OT-CycleGAN uses lesser number of parameters than the proposed model, it cannot capture additional information to balance information asymmetry between two image domains, and eventually fails to beat the proposed model. Through the qualitative representation presented in Fig. 6.4, it

Table 6.3: Statistical significance analysis of different algorithms with respect to Wilcoxon signed-rank (one-tailed) test on TCGA Data set

Methods	Dice	Jaccard	Precision	Recall
Ostrich \R	2.37E-23	2.34E-23	5.57E-07	2.01E-09
CycleGAN	9.05E-27	6.08E-27	1.66E-30	7.70E-10
OT-CycleGAN	8.58E-25	6.83E-25	7.49E-04	6.37E-18
Asym-CycleGAN	*8.56E-02*	*7.80E-02*	1.13E-03	*7.67E-02*
Checkerboard	7.74E-13	2.49E-13	8.57E-33	1.21E-02
U-Net	8.61E-32	7.06E-32	1.89E-33	4.45E-02
Mask-R-CNN	2.74E-28	3.20E-28	1.55E-03	1.41E-20
U-Net++	9.70E-17	6.71E-17	2.86E-07	1.88E-03
HoVer-Net	1.63E-30	1.23E-30	*6.85E-02*	4.13E-32
MoNS	1.46E-13	3.65E-14	2.11E-33	*1.10E-01*
Stardist	7.73E-22	6.91E-22	*3.40E-01*	3.03E-23
Swin-MIL	9.15E-32	8.48E-32	3.83E-02	1.47E-32
BoNuS	*8.18E-02*	*5.86E-02*	7.36E-06	*9.00E-01*

Table 6.4: Number of parameters involved in each of the baseline image-to-image translation models (in millions)

Methods	No. of Parameters (M)
CycleGAN	18.035
OT-CycleGAN	4.353
Asym-CycleGAN	27.053
Ostrich	7.271

can also be observed that the proposed model performs better than all the aforementioned baseline models.

6.3.2.3 Importance of Spatially-Constrained Squeeze Operation

Table 6.5: Performance in binary semantic segmentation

Model	Dice	Jaccard	Precision	Recall
Ostrich	0.810099	0.746830	0.817761	0.802580

The performance of the proposed Ostrich model with spatially-constrained squeeze operation is compared with that of the same model with checkerboard masking based squeeze operation, explained in [124], and the corresponding results are reported in Table 6.1. The values reported in Table 6.1 ensure the fact that spatially-constrained squeeze operation based proposed Ostrich model outperforms the model with checkerboard masking based squeeze operation, with respect to all the segmentation evaluation indices. The qualitative comparison of the performance of the proposed model and checkerboard masking based model is presented in Fig. 6.5. It can be observed in Fig. 6.5 that the proposed Ostrich model performs better than the checkerboard masking based model in separating overlapping nuclei structures. This is due to the fact that the proposed model can capture
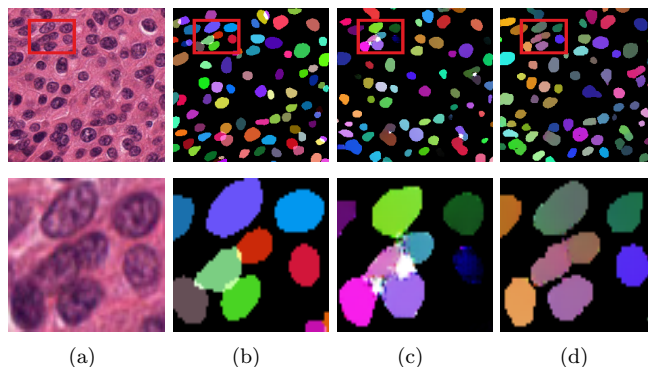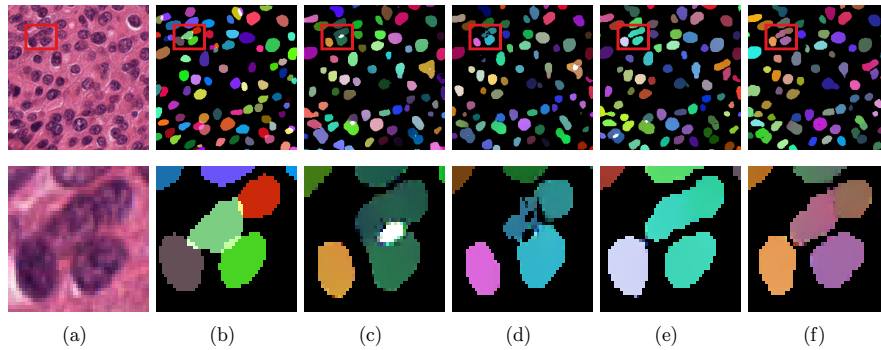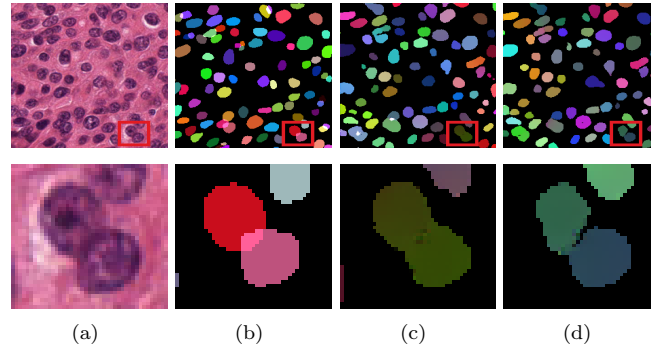
Figure 6.5: Performance analysis of nuclei segmentation: (a) Original histological image, (b) Ground-truth; Segmentation performance of (c) checkerboard masking based squeeze operation and (d) spatially-constrained squeeze operation. Top row presents marked-up images from TCGA data, and bottom row presents the zoomed-in region from the marked-up patches.
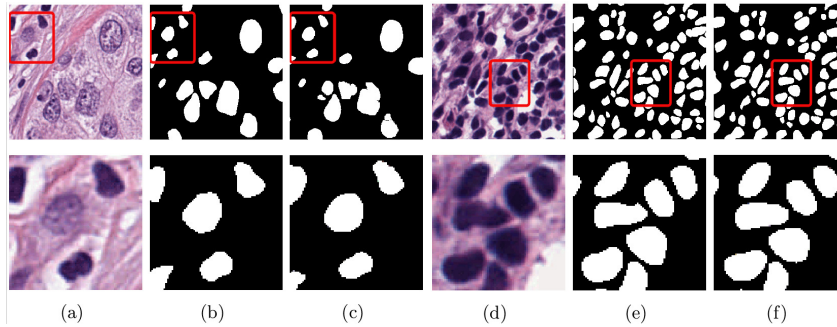
Figure 6.6: Performance analysis in binary semantic segmentation: (a) histological image patch 1, (b) ground-truth segmentation map 1, (c) predicted map 1 ; (d) image patch 2, (e) ground-truth segmentation 2 and (f) predicted map 2. Top row presents marked-up images from TCGA data, and bottom row presents the zoomed-in region from the marked-up patches.

spatial connectivity through spatially-constrained squeeze operation, which checkerboard masking based model fails to capture. The p-values reported in Table 6.2 and 6.3 establish the fact that proposed Ostrich model with spatially-constrained squeeze operation performs significantly better than the checkerboard masking based counterpart.

6.3.2.4 Performance in Binary Semantic Segmentation

The performance of the proposed Ostrich model on TCGA data set in binary semantic segmentation is validated through the qualitative results presented in Fig. 6.6. It presents two image patches, their corresponding ground-truth images and two binary segmentation maps predicted by the proposed Ostrich model. From the qualitative analysis, it is clear that the proposed model can capture heterogeneous nuclei shapes. It is possible due to the fact that the proposed model does not assume the shape of the nuclei objects apriori. The results reported in Table 6.5 also ensures the fact that the proposed model performs well

Figure 6.7: (a) Original images of TCGA data set, (b) ground-truth segmentation map; and segmentation maps obtained using existing nuclei segmentation methods: (c) U-Net, (d) Mask-R-CNN and (e) U-Net++.

in binary semantic segmentation.

Table 6.6: Quantitative performance analysis in nuclei segmentation on CoNIC Data set: State-of-the-art models vs Ostrich

Methods	Dice	Jaccard	Precision	Recall
Ostrich	**0.741107**	**0.591310**	**0.769381**	**0.793093**
U-Net	0.715636	0.561178	0.704462	0.681140
Mask-R-CNN	0.717821	0.562671	0.725431	0.706849
U-Net++	0.721316	0.568022	0.721591	0.706789
HoVer-Net	0.740318	0.590560	0.747550	0.698034
MoNS	0.722989	0.568692	0.722632	0.733005
Stardist	0.374534	0.240549	0.716413	0.275528
Swin-MIL	0.715991	0.560825	0.764054	0.682995
BoNuS	0.724678	0.571128	0.753906	0.768233

6.3.2.5 Comparison with State-of-the-Art Methods

The performance of the proposed Ostrich model is finally compared with that of several state-of-the-art nuclei segmentation approaches on TCGA data set and the results with respect to different quantitative indices are reported in Table 6.1. From the results reported in Table 6.1, it can be observed that, in case of TCGA data, the proposed model outperforms all the state-of-the-art nuclei segmentation approaches, with respect to three evaluation indices, namely, Dice, Jaccard and precision. With respect to recall score, BoNuS performs slightly better than the proposed Ostrich model. The qualitative performance of the

Figure 6.8: (a) Original images of TCGA data set, (b) ground-truth segmentation map; and segmentation maps obtained using existing nuclei segmentation methods: (c) HoVer-Net, (d) MoNS and (e) Stardist.

Table 6.7: Statistical significance analysis of different algorithms with respect to Paired-t (one-tailed) test on CoNIC data

Methods	Dice	Jaccard	Precision	Recall
U-Net	1.21E-85	1.27E-86	9.99E-99	9.99E-99
Mask-R-CNN	9.99E-99	1.11E-99	9.99E-99	9.99E-99
U-Net++	1.02E-63	2.46E-87	9.99E-99	9.99E-99
HoVer-Net	*8.81E-02*	*1.44E-01*	4.57E-98	9.99E-99
MoNS	9.99E-99	9.99E-99	9.99E-99	9.99E-99
Stardist	9.99E-99	1.11E-99	1.24E-82	9.99E-99
Swin-MIL	9.99E-99	9.99E-99	5.75E-20	9.99E-99
BoNuS	4.28E-78	1.12E-93	9.99E-99	9.99E-99

Ostrich model, along with comparisons with state-of-the-art approaches, in nuclei segmentation, are presented in Fig. 6.7, Fig. 6.8 and Fig. 6.9. using two image patches from the TCGA data set. Analyzing the p-values reported in Table 6.2, it can be observed that, with respect paired-t test, the proposed Ostrich model performs significantly better in 29 out of 36 cases, better but not significantly in 6 cases. The p-values reported in Table 6.3 also ensure that the proposed model performs significantly better in 29 out of 36 cases, better but not significantly in 6 cases, with respect Wilcoxon signed-rank test.

6.3.3 Performance Analysis on CoNIC Data

Finally, the performance of the proposed Ostrich model is compared with that of existing nuclei segmentation methods on the CoNIC data set and the corresponding results are reported in Table 6.6. From the results provided in Table 6.6, it can be easily observed that,
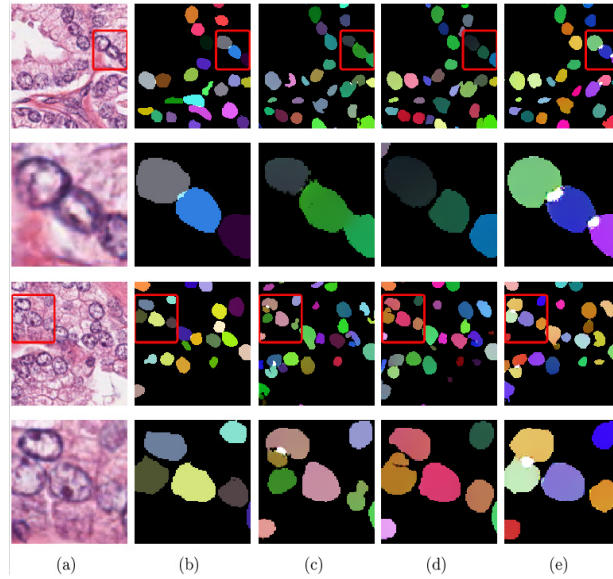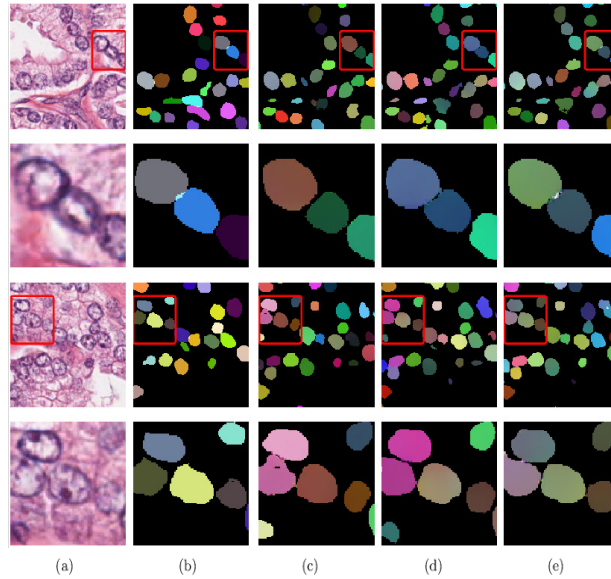
Figure 6.9: (a) Original images of TCGA data set, (b) ground-truth segmentation map; and segmentation maps obtained using existing nuclei segmentation methods: (c) Swin-MIL, (d) BoNuS and (e) Ostrich.

Table 6.8: Statistical significance analysis of different algorithms with respect to Wilcoxon signed-rank (one-tailed) test on CoNIC data

Methods	Dice	Jaccard	Precision	Recall
U-Net	1.17E-71	2.67E-71	9.99E-99	9.99E-99
Mask-R-CNN	9.99E-99	9.99E-99	9.99E-99	9.99E-99
U-Net++	9.99E-99	9.99E-99	9.99E-99	9.99E-99
HoVer-Net	1.93E-01	2.56E-01	9.99E-99	9.99E-99
MoNS	9.99E-99	9.99E-99	9.99E-99	9.99E-99
Stardist	9.99E-99	1.11E-99	1.08E-83	9.99E-99
Swin-MIL	9.99E-99	9.99E-99	2.37E-31	9.99E-99
BoNuS	9.99E-99	9.99E-99	9.99E-99	9.99E-99

in case of CoNIC data set, the proposed Ostrich model outperforms existing nuclei segmentation approaches with respect all the evaluation indices. The high precision and low recall value of Stardist ensures the fact that it is confident about a few positive instances it predicts, but misses a substantial number of true positive instances. The segmentation performance of the proposed Ostrich model is also validated through a qualitative comparison presented in Fig. 6.10. The qualitative comparison through zoomed-in regions of the segmentation maps makes it clear that the proposed Ostrich model captures the nuclei structures better than most of the existing approaches. Finally, analyzing the p-values reported in Table 6.7, it can be noted that, the proposed Ostrich model performs significantly better in 34 out of 36 cases, better but not significantly in 2 cases, with respect paired-t test. From the p-values reported in Table 6.8, it is evident that the proposed model performs significantly better in 34 out of 36 cases, better but not significantly in 2 cases, with respect Wilcoxon signed-rank test. In Table 6.7 and 6.8, the p-values less than
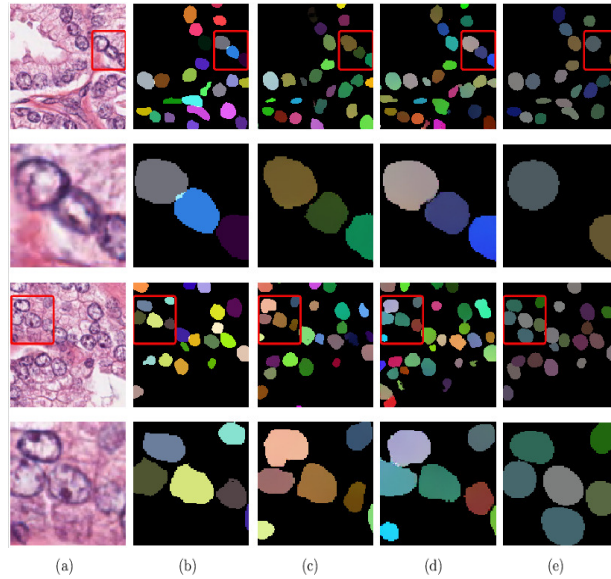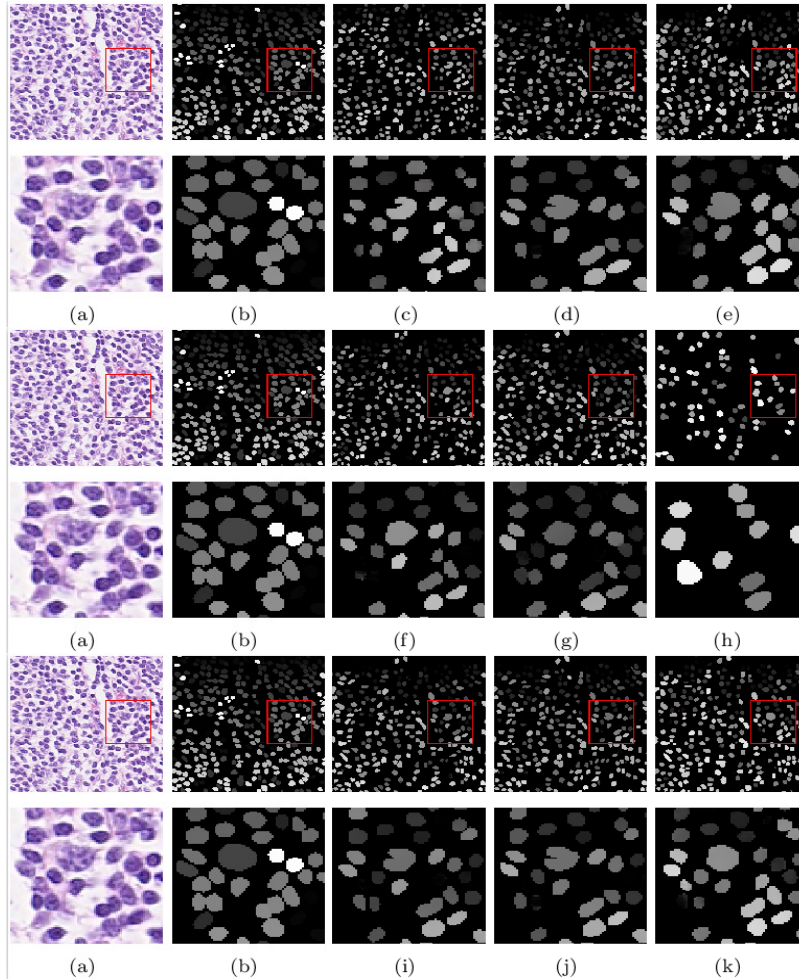
Figure 6.10: (a) Original images of CoNIC data set, (b) ground-truth segmentation map; and segmentation maps obtained using different nuclei segmentation methods: (c) U-Net, (d) Mask-R-CNN, (e) U-Net++, (f) HoVer-Net, (g) MoNS, (h) Stardist, (i) Swin-MIL, (j) BoNuS and (k) Ostrich.

or equal to $9.99E - 99$ are reported as $9.99E - 99$.

The quantitative results reported in Table 6.1 and 6.6, and the qualitative analyses presented in Fig. 6.7-6.9 and Fig. 6.10 ensure the fact that the proposed Ostrich model performs better than the state-of-the-art nuclei segmentation approaches in almost all the cases. The proposed model performs consistently well in different types of nuclei segmentation tasks due to the following facts: (a) it employs image-to-image translation paradigm, which is invariant to different target domain representations, (b) incorporation of an additional embedding space helps the proposed model to balance information asymmetry between two image domains.

6.4   Conclusion

Nuclei segmentation is one of the fundamental and critical steps in histological image analysis as it aids in subsequent analysis of tissue specimens to estimate tumor cellularity, which is very crucial in the diagnosis and prognosis of cancer patients. In this respect, the main contributions of the current study are listed as follows:

- development of a deep generative model for segmenting nuclear regions from histological images, which considers an embedding space to handle information-imbalance between two image domains;

- development of an invertible generator, which achieves asymmetric image-to-image translation with lower network complexity;

- introduction of a regularization term to perturb image generation towards the accurate direction at the initial stage of training; and

- demonstrating the effectiveness of the proposed model, along with a comparison with existing algorithms, on publicly available H&E stained histology image data sets.

The proposed model utilizes image-to-image translation paradigm, which enhances the generalizability and applicability of the model in different types of nuclei segmentation tasks. From the quantitative and qualitative analysis, it is evident that the proposed model performs consistently well in both binary semantic segmentation and instance segmentation task with different types of target domain image representations. The proposed model does not consider any prior information about the shape of nuclei structures, which helps the model to capture heterogeneous nuclei shapes from tissue images belonging to different human organs.

Stain color normalization is one of the primary and necessary steps in histological image analysis, which enables better nuclei segmentation from histological images. On the other hand, accurate nuclei segmentation makes color normalization rather trivial. Hence, these two intertwined procedures are integrated so that they take advantage from each other, simultaneously yielding better nuclei segmentation as well as better color normalization. In this respect, the next chapter introduces a new deep generative model for simultaneous nuclei segmentation and color normalization of histological images.

Chapter 7

Simultaneous Nuclei Segmentation and Color Normalization of Histological Images

7.1 Introduction

The most important characteristic of histological images is that they contain an enormous amount of phenotypic information, which makes computer-aided diagnosis more accurate than other imaging modalities. Nuclei segmentation, as described in Chapter 6, is one of the crucial steps in histological image analysis as it presents important morphological information, which helps in therapeutic diagnosis and prognosis of cancer. Along with the challenges associated with nuclei segmentation from histological images, such as heterogeneous morphologies, color, texture, overlapped or clustered nuclei, another important factor which degrades the performance of nuclei segmentation is inadmissible inter and intra-specimen disagreement in the appearance of stained tissue color. This variability in color appearance hinders accurate nuclei segmentation as it affects the contrast between nuclei structures, other cellular components and surrounding cytoplasm.

It has been highlighted by an experimental study in Chapter 5 that color normalization is required as a critical preprocessing step, which enables better nuclei segmentation from histological images. On the other hand, accurate nuclei segmentation makes color normalization rather trivial. Hence, color normalization and nuclei segmentation can be considered as two intertwined procedures. These two procedures can be integrated together so that they can take advantage from each other, simultaneous yielding better nuclei segmentation and color normalization. Although some works [16], [87], [100] have been proposed that address the nuclei segmentation problem followed by stain color normalization, the simultaneous nuclei segmentation and color normalization problem is not addressed in the literature. In [114], the authors have proposed a simultaneous nuclei segmentation and classification model, but they did not address the stain color normalization problem.

In this background, this chapter introduces a simultaneous nuclei segmentation and color normalization model (SNSCN), that integrates the merits of spatial attention and

truncated normal distribution. The SNSCN model assumes that the latent color appearance code is pairwise independent of nuclei segmentation map and embedding map, which embeds other tissue-level details. To deal with overlapping nature of histochemical stains, the proposed model assumes that the latent color appearance code is sampled from a mixture of truncated normal distributions. The concept of spatial attention is incorporated to segment nuclei regions from histological images. The effectiveness of the proposed model in nuclei segmentation as well as stain color normalization is demonstrated on publicly available histological image data sets containing H&E stained images.

The rest of this chapter is organized as follows: Section 7.2 describes the proposed SNSCN model along with the generative module (Section 7.2.3.1) and reconstructive module (Section 7.2.3.2). The proficiency of the SNSCN framework is analyzed with reference to several state-of-the-art nuclei segmentation and color normalization approaches on H&E stained histological image data sets in Section 7.3. Concluding remarks are provided in Section 7.4.

7.2 SNSCN: Proposed Model

This section presents a deep generative model, termed as SNSCN, for simultaneous nuclei segmentation and color normalization of histological images.

7.2.1 Problem Statement

Given a histological image set $\mathcal{X}$, containing a set of n number of images $\{x_i : i = 1, 2, \cdots, n\}$, the goal is to build a model that, in the first stage, takes a non-normalized histological image $x_i \in \mathcal{X} \subset \mathbb{R}^{H \times W \times 3}$ as input and outputs corresponding nuclei segmentation map y_i. The set of these output maps $\{y_i\}$ forms segmentation map domain $\mathcal{Y}$, that is, $y_i \in \mathcal{Y} \subset \mathbb{R}^{H \times W \times 3}$ (in case of instance segmentation) or $y_i \in \mathcal{Y} \subset \mathbb{R}^{H \times W}$ (in case of binary semantic segmentation). In the second stage, the model takes the nuclei segmentation map y_i as input and outputs the color normalized image corresponding to non-normalized input image x_i. As described in Chapter 6, there is an information-disparity between the two image domains: information-rich histological image space $\mathcal{X}$ and information-poor segmentation map domain $\mathcal{Y}$. Any arbitrary pair of images x_i and x_j, $i \neq j$, within histological image set $\mathcal{X}$, exhibit color variation due to different factors associated with color disagreement. Hence, the aim of the current study is to develop a deep generative framework that, with proper training, simultaneously will be able to generate color normalized histological image corresponding to each non-normalized histological slide image, and segmentation annotation map corresponding to the normalized image by handling the information-asymmetry present between two image domains.

7.2.2 Model Structure

Different components of the proposed SNSCN model is shown in the block diagram presented in Fig. 7.1. The proposed model consists of five deep networks: a color appearance encoder $\mathcal{E}_c$, which extracts the color appearance information, a segmentation map generator $\mathcal{F}_\phi$, which generates nuclei segmentation map corresponding to a histological image,
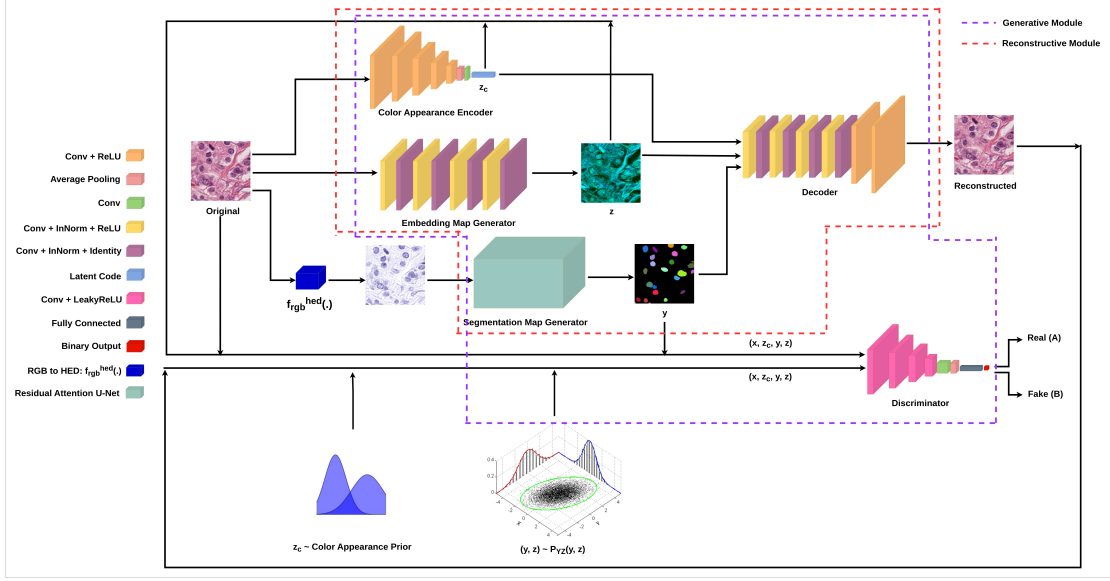
Figure 7.1: Block diagram of the proposed deep generative model SNSCN for simultaneous nuclei segmentation and color normalization of histological images. In the diagram, 'Conv' means convolutional layer, 'ReLU' denotes rectified linear unit, 'InNorm' represents instance normalization, 'Identity' means identity function, i.e., $f(x) = x$, 'LeakyReLU' denotes leaky rectified linear unit.

an embedding map generator $\mathcal{E}_\omega$, that encodes additional information while translating from information-rich image space to information-poor segmentation map domain, a decoder $\mathcal{G}$ and a discriminator $\mathcal{D}$. Each training set image x is fed simultaneously into the three networks: the color appearance encoder $\mathcal{E}_c(x; \theta_{\mathcal{E}_c})$, the segmentation map generator $\mathcal{F}_\phi(x; \theta_{\mathcal{F}_\phi})$ and embedding map generator $\mathcal{E}_\omega(x; \theta_{\mathcal{E}_\omega})$, which eventually output the latent color appearance code z_c, nuclei segmentation map y and embedding information z, respectively. Latent color appearance code z_c, along with nuclei segmentation map y and embedding information z, are then fed into the decoder $\mathcal{G}(z_c, y, z; \theta_{\mathcal{G}})$ as inputs, which generates the reconstructed color normalized image $\hat{x}$. The discriminator $\mathcal{D}(x, z_c, y, z; \theta_{\mathcal{D}})$ takes inputs in the form of quadruplets (x, z_c, y, z) and discriminates real encoding, where x comes from real data distribution, z_c, y and z correspond to latent code obtained through $\mathcal{E}_c$, segmentation map generated through $\mathcal{F}_\phi$, and embedding map generated through $\mathcal{E}_\omega$, respectively, from generated/fake encoding, where x denotes the reconstructed sample $\hat{x}$ and z_c, y, z come from corresponding prior distributions. Here, $\theta_{\mathcal{E}_c}$, $\theta_{\mathcal{F}_\phi}$, $\theta_{\mathcal{E}_\omega}$, $\theta_{\mathcal{G}}$ and $\theta_{\mathcal{D}}$ represent the network parameters associated with the deep networks $\mathcal{E}_c$, $\mathcal{F}_\phi$, $\mathcal{E}_\omega$, $\mathcal{G}$ and $\mathcal{D}$, respectively. The networks $\mathcal{E}_c$, $\mathcal{F}_\phi$, $\mathcal{E}_\omega$, $\mathcal{G}$ and $\mathcal{D}$ should be non-linear differentiable functions so that the error values can be back-propagated during the training phase. Nuclei regions in a Hematoxylin and Eosin (H&E)-stained histological image are generally highlighted by hematoxylin (H) stain. The H-channel of an image $x \in \mathcal{X}$ is extracted through $f_{rgb}^{hed}(x)[:,:,0]$, where $f_{rgb}^{hed}(.)$ represents a function that takes RGB image as input and outputs corresponding image in Hematoxylin-Eosin-DAB (HED) colorspace. Hence, a spatial attention operation through deep model $\mathcal{F}_\phi$ is applied on H-channel corresponding to each

125

histological image to extract nuclei regions. To incorporate spatial attention mechanism in the network, existing residual attention U-Net model [109] is used in this study.

7.2.3 Objective Function

The objective of the model is to extract latent color appearance information, which is pairwise independent of nuclei segmentation map and tissue-level embedding information, and to learn the fake encoding as close as possible to the real encoding by minimizing the following objective function:

$$J_{\text{total}} = \lambda_{adv} \times J_{\text{adv}} + \lambda_{rec} \times J_{\text{rec}}, \tag{7.1}$$

where the minimization objective terms J_{adv} and J_{rec} are attributed by the generative module and the reconstructive module, respectively, which is demonstrated next. Here, λ_{adv} and λ_{rec} denote the relative importance of the adversarial objective J_{adv} and the reconstruction objective J_{rec}, respectively. In this current study, J_{total} is considered to be the convex combination of J_{adv} and J_{rec}, that is, $\lambda_{adv} + \lambda_{rec} = 1$.

7.2.3.1 Generative Module

The generative module is comprised of all five deep networks: $\mathcal{E}_c$, $\mathcal{F}_\phi$, $\mathcal{E}_\omega$, $\mathcal{G}$ and $\mathcal{D}$, as marked in Fig. 7.1 with violet dashed line. The generative module incorporates the assumption that the latent color appearance information is pairwise independent of nuclei segmentation map and embedding information. The objective functions, corresponding to the generative module, are given as follows:

$$J_{\text{G}}(\mathcal{D}) = \min_{\mathcal{D}} J_1(\mathcal{E}_c, \mathcal{F}_\phi, \mathcal{E}_\omega, \mathcal{G}, \mathcal{D}), \quad \text{where} \tag{7.2}$$

$$J_1(\mathcal{E}_c, \mathcal{F}_\phi, \mathcal{E}_\omega, \mathcal{G}, \mathcal{D}) = \underbrace{E_{x \sim P_\mathcal{X}(x)} E_{z_c \sim P_{\mathcal{E}_c}(z_c|x)} E_{y \sim P_{\mathcal{F}_\phi}(y|x)} E_{z \sim P_{\mathcal{E}_\omega}(z|x)} (A - \mathcal{D}[x, z_c, y, z])^2}_{R} +$$

$$\underbrace{E_{z_c \sim P_{\mathcal{Z}_c}(z_c)} E_{y \sim P_\mathcal{Y}(y)} E_{z \sim P_\mathcal{Z}(z)} E_{x \sim P_\mathcal{G}(x|z_c,y,z)} (B - \mathcal{D}[x, z_c, y, z])^2}_{G},$$

Here, A and B denote the labels, which are assigned by discriminator $\mathcal{D}$, to designate real and generated/fake encoding, respectively, and

$$J_{\text{G}}(\mathcal{G}) = \min_{\mathcal{G}} J_2(\mathcal{E}_c, \mathcal{F}_\phi, \mathcal{E}_\omega, \mathcal{G}, \mathcal{D}), \quad \text{where}$$

$$J_2(\mathcal{E}_c, \mathcal{F}_\phi, \mathcal{E}_\omega, \mathcal{G}, \mathcal{D}) = E_{z_c \sim P_{\mathcal{Z}_c}(z_c)} E_{y \sim P_\mathcal{Y}(y)} E_{z \sim P_\mathcal{Z}(z)} E_{x \sim P_\mathcal{G}(x|z_c,y,z)} (C - \mathcal{D}[x, z_c, y, z])^2 \tag{7.3}$$

C denotes the label, assigned by discriminator $\mathcal{D}$ to designate fake encoding, as desired by decoder $\mathcal{G}$.

In part R of (7.2), $P_\mathcal{X}(x)$ represents the real data distribution, $P_{\mathcal{E}_c}(z_c \mid x)$, $P_{\mathcal{F}_\phi}(y \mid x)$ and $P_{\mathcal{E}_\omega}(z \mid x)$ represent the conditional distributions corresponding to the encoder $\mathcal{E}_c$, and

generators $\mathcal{F}_\phi$, $\mathcal{E}_\omega$, respectively, given the data sample x. In part G of (7.2), $P_{\mathcal{Z}_c}(z_c)$, $P_{\mathcal{Y}}(y)$ and $P_{\mathcal{Z}}(z)$ denote the prior distributions corresponding to latent color appearance code z_c, nuclei segmentation map y and embedding information z, respectively, and $P_{\mathcal{G}}(x \mid z_c, y, z)$ represents the conditional distribution corresponding to generated sample $\hat{x}$, given the latent code z_c, and generated maps y and z. Now, (7.2) can be simplified as follows:

$$J_1(\mathcal{E}_c, \mathcal{F}_\phi, \mathcal{E}_\omega, \mathcal{G}, \mathcal{D}) = \int_x P_{\mathcal{X}}(x) \int_{z_c} P_{\mathcal{E}_c}(z_c \mid x) \int_y P_{\mathcal{F}_\phi}(y \mid x) \int_z P_{\mathcal{E}_\omega}(z \mid x)$$

$$\times (A - \mathcal{D}[x, z_c, y, z])^2 dz\,dy\,dz_c\,dx$$

$$+ \int_{z_c} P_{\mathcal{Z}_c}(z_c) \int_y P_{\mathcal{Y}}(y) \int_z P_{\mathcal{Z}}(z) \int_x P_{\mathcal{G}}(x \mid z_c, y, z)(B - \mathcal{D}[x, z_c, y, z])^2 dx\,dz\,dy\,dz_c$$

$$= \int_{\{x, z_c, y, z\}} [P_{\mathcal{X}}(x) P_{\mathcal{E}_c}(z_c \mid x) P_{\mathcal{F}_\phi}(y \mid x) P_{\mathcal{E}_\omega}(z \mid x)(A - \mathcal{D}[x, z_c, y, z])^2$$

$$+ P_{\mathcal{Z}_c}(z_c) P_{\mathcal{Y}}(y) P_{\mathcal{Z}}(z) P_{\mathcal{G}}(x \mid z_c, y, z)(B - \mathcal{D}[x, z_c, y, z])^2] dx\,dz_c\,dy\,dz$$

$$= \int_{\{x, z_c, y, z\}} [P_{\mathcal{X}}(x) P_{\mathcal{E}_c \mathcal{F}_\phi \mathcal{E}_\omega}(z_c, y, z \mid x)(A - \mathcal{D}[x, z_c, y, z])^2 +$$

$$P_{\mathcal{Z}_c \mathcal{Y} \mathcal{Z}}(z_c, y, z) P_{\mathcal{G}}(x \mid z_c, y, z)(B - \mathcal{D}[x, z_c, y, z])^2] dx\,dz_c\,dy\,dz$$

as latent code z_c is assumed to be pairwise independent of segmentation map y and embedding information z. So,

$$J_1(\cdot) = \int_{\{x, z_c, y, z\}} [P_{\mathcal{E}_c \mathcal{F}_\phi \mathcal{E}_\omega \mathcal{X}}(x, z_c, y, z)(A - \mathcal{D}[x, z_c, y, z])^2 +$$

$$P_{\mathcal{G} \mathcal{Z}_c \mathcal{Y} \mathcal{Z}}(x, z_c, y, z)(B - \mathcal{D}[x, z_c, y, z])^2] dx\,dz_c\,dy\,dz. \tag{7.4}$$

At first, the optimal discriminator $\mathcal{D}$ is to be computed for any given encoder-generator-decoder combination $(\mathcal{E}_c, \mathcal{F}_\phi, \mathcal{E}_\omega, \mathcal{G})$.

Proposition 7.1. *For a fixed combination of color appearance encoder $\mathcal{E}_c$, nuclei segmentation map generator $\mathcal{F}_\phi$, embedding map generator $\mathcal{E}_\omega$ and decoder $\mathcal{G}$, the optimal discriminator $\mathcal{D}$ can be computed as:*

$$\mathcal{D}^*[x, z_c, y, z] = \frac{A.P_{\mathcal{E}_c \mathcal{F}_\phi \mathcal{E}_\omega \mathcal{X}}(x, z_c, y, z) + B.P_{\mathcal{G} \mathcal{Z}_c \mathcal{Y} \mathcal{Z}}(x, z_c, y, z)}{P_{\mathcal{E}_c \mathcal{F}_\phi \mathcal{E}_\omega \mathcal{X}}(x, z_c, y, z) + P_{\mathcal{G} \mathcal{Z}_c \mathcal{Y} \mathcal{Z}}(x, z_c, y, z)} \tag{7.5}$$

Proof.
For a fixed combination of $\mathcal{E}_c$, $\mathcal{E}_s$ and $\mathcal{G}$, the learning criterion for discriminator $\mathcal{D}$ is to

minimize $J_1(\mathcal{E}_c, \mathcal{E}_s, \mathcal{G}, \mathcal{D})$ with respect to $\mathcal{D}$ by computing partial derivative as follows:

$$\frac{\partial J_1(\mathcal{E}_c, \mathcal{E}_s, \mathcal{G}, \mathcal{D})}{\partial \mathcal{D}[x, z_c, z_s]} = 2P_{\mathcal{E}_c\mathcal{E}_s\mathcal{X}}(x, z_c, z_s)(\mathcal{D}[x, z_c, z_s]$$
$$-A) + 2P_{\mathcal{G}\mathcal{Z}_c\mathcal{Z}_s}(x, z_c, z_s)(\mathcal{D}[x, z_c, z_s] - B) \tag{7.6}$$

Let the optimal discriminator be denoted as $\mathcal{D}^*[x, z_c, z_s]$. So, at optimal point corresponding to the discriminator $\mathcal{D}$:

$$\left.\frac{\partial J_1(\mathcal{E}_c, \mathcal{E}_s, \mathcal{G}, \mathcal{D})}{\partial \mathcal{D}[x, z_c, z_s]}\right|_{\mathcal{D}=\mathcal{D}^*} = 0;$$

$$\Rightarrow 2\{P_{\mathcal{E}_c\mathcal{E}_s\mathcal{X}}(x, z_c, z_s)(\mathcal{D}^*[x, z_c, z_s] - A)+$$
$$P_{\mathcal{G}\mathcal{Z}_c\mathcal{Z}_s}(x, z_c, z_s)(\mathcal{D}^*[x, z_c, z_s] - B)\} = 0; \quad [\text{using (7.6)}]$$

$$\Rightarrow \mathcal{D}^*[x, z_c, z_s] = \frac{\begin{array}{c}A.P_{\mathcal{E}_c\mathcal{E}_s\mathcal{X}}(x, z_c, z_s)+\\ B.P_{\mathcal{G}\mathcal{Z}_c\mathcal{Z}_s}(x, z_c, z_s)\end{array}}{P_{\mathcal{E}_c\mathcal{E}_s\mathcal{X}}(x, z_c, z_s) + P_{\mathcal{G}\mathcal{Z}_c\mathcal{Z}_s}(x, z_c, z_s)} \tag{7.7}$$

Theoretically, for optimal combination of encoder-generator-decoder $(\mathcal{E}_c, \mathcal{F}_\phi, \mathcal{E}_\omega, \mathcal{G})^*$, the discriminator should not be able to distinguish real encoding from fake encoding as decoder $\mathcal{G}$ learns the intrinsic real joint distribution. As a result, at equilibrium point, the optimal discriminator $\mathcal{D}^*$ cannot discriminate between real and generated/fake encoding, and designates both of them by the same label C that the decoder $\mathcal{G}$ wants $\mathcal{D}$ to believe. So, from (7.4), it can be written that

$$(\mathcal{E}_c, \mathcal{F}_\phi, \mathcal{E}_\omega, \mathcal{G})^* = \arg\min_{\mathcal{E}_c, \mathcal{F}_\phi, \mathcal{E}_\omega, \mathcal{G}} J_2(\mathcal{E}_c, \mathcal{F}_\phi, \mathcal{E}_\omega, \mathcal{G}, \mathcal{D}^*), \quad \text{where}$$

$$J_2(\mathcal{E}_c, \mathcal{F}_\phi, \mathcal{E}_\omega, \mathcal{G}, \mathcal{D}^*) = E_{x\sim P_\mathcal{X}(x)}E_{z_c\sim P_{\mathcal{E}_c}(z_c|x)}E_{y\sim P_{\mathcal{F}_\phi}(y|x)}E_{z\sim P_{\mathcal{E}_\omega}(z|x)}(C - \mathcal{D}^*[x, z_c, y, z])^2$$

$$+ E_{z_c\sim P_{\mathcal{Z}_c}(z_c)}E_{y\sim P_\mathcal{Y}(y)}E_{z\sim P_\mathcal{Z}(z)}E_{x\sim P_\mathcal{G}(x|z_c,y,z)}(C - \mathcal{D}^*[x, z_c, y, z])^2$$

[adding terms independent of $\mathcal{G}$ does not impact optimization]

$$= \int_{\{x,z_c,y,z\}} \{P_{\mathcal{E}_c\mathcal{F}_\phi\mathcal{E}_\omega\mathcal{X}}(x, z_c, y, z)(C-\mathcal{D}^*[x, z_c, y, z])^2 + P_{\mathcal{G}\mathcal{Z}_c\mathcal{Y}\mathcal{Z}}(x, z_c, y, z)(C-\mathcal{D}^*[x, z_c, y, z])^2\}$$

$$dxdz_cdydz$$

$$= \int_{\{x,z_c,y,z\}} \{P_{\mathcal{E}_c\mathcal{F}_\phi\mathcal{E}_\omega\mathcal{X}}(x, z_c, y, z) + P_{\mathcal{G}\mathcal{Z}_c\mathcal{Y}\mathcal{Z}}(x, z_c, y, z)\}$$

$$\times \left[C - \frac{A.P_{\mathcal{E}_c\mathcal{F}_\phi\mathcal{E}_\omega\mathcal{X}}(x, z_c, y, z) + B.P_{\mathcal{G}\mathcal{Z}_c\mathcal{Y}\mathcal{Z}}(x, z_c, y, z)}{P_{\mathcal{E}_c\mathcal{F}_\phi\mathcal{E}_\omega\mathcal{X}}(x, z_c, y, z) + P_{\mathcal{G}\mathcal{Z}_c\mathcal{Y}\mathcal{Z}}(x, z_c, y, z)}\right]^2 dxdz_cdydz.$$

Hence, the adversarial objective term, attributed by the generative module, can be formulated as follows:

$$J_{\text{adv}} = J_{\text{G}}(\mathcal{D}) + J_{\text{G}}(\mathcal{G}). \tag{7.8}$$

7.2.3.2 Reconstructive Module

The reconstructive module consists of four deep networks: $\mathcal{E}_c$, $\mathcal{F}_\phi$, $\mathcal{E}_\omega$ and $\mathcal{G}$, as marked in Fig. 7.1 with red dashed line. The reconstructive module also assumes that the latent color appearance code z_c is pairwise independent of generated nuclei segmentation map y and embedding information z. Now, the joint distribution $P_{\mathcal{G}\mathcal{Z}_c\mathcal{Y}\mathcal{Z}}$, which is associated with the reconstruction of normalized image samples, can be decomposed as follows:

$$P_{\mathcal{G}\mathcal{Z}_c\mathcal{Y}\mathcal{Z}}(x, z_c, y, z) = P_{\mathcal{G}}(x \mid z_c, y, z) P_{\mathcal{Z}_c\mathcal{Y}\mathcal{Z}}(z_c, y, z). \tag{7.9}$$

Now, z_c is assumed to be pairwise independent of y and z. But, nuclei segmentation map y and embedding information z capture complementary information from histological images. So, y and z cannot be assumed to be independent of each other. Hence,

$$P_{\mathcal{G}\mathcal{Z}_c\mathcal{Y}\mathcal{Z}}(x, z_c, y, z) = P_{\mathcal{G}}(x \mid z_c, y, z) P_{\mathcal{Z}_c}(z_c) P_{\mathcal{Y}\mathcal{Z}}(y, z). \tag{7.10}$$

The inference problem is to compute the joint conditional density of latent code z_c, and generated maps y and z, given the real histological image sample x. So,

$$P(z_c, y, z \mid x) = \frac{P_{\mathcal{G}\mathcal{Z}_c\mathcal{Y}\mathcal{Z}}(x, z_c, y, z)}{P(x)}, \tag{7.11}$$

where $P(x)$ denotes the marginal likelihood or model evidence, which can be computed as follows:

$$P(x) = \int\limits_{z_c} \int\limits_{y} \int\limits_{z} P_{\mathcal{G}\mathcal{Z}_c\mathcal{Y}\mathcal{Z}}(x, z_c, y, z) dz_c dy dz. \tag{7.12}$$

The computation of model evidence is typically intractable due to the fact that the integral in (7.12) requires integration with respect to multi-dimensional latent representation z_c, and generated nuclei segmentation map y and embedding information z. So, an auxiliary approximate distribution $Q(z_c, y, z)$, defined in such a way that it has closed form solution, must be utilized to approximate the inference $P(z_c, y, z \mid x)$ by minimizing the Kullback-Leibler (KL) divergence between $P(z_c, y, z \mid x)$ and $Q(z_c, y, z)$ as follows:

$$D_{KL}[Q(z_c, y, z) \mid\mid P(z_c, y, z \mid x)] = -\int\limits_{z_c} \int\limits_{y} \int\limits_{z} Q(z_c, y, z) \log\left[\frac{P(z_c, y, z \mid x)}{Q(z_c, y, z)}\right] dz_c dy dz$$

$$= -E_{Q(z_c, y, z)}\left[\log\left(\frac{P_{\mathcal{G}\mathcal{Z}_c\mathcal{Y}\mathcal{Z}}(x, z_c, y, z)}{Q(z_c, y, z)}\right)\right] + \log P(x). \tag{7.13}$$

Now, the minimum value of KL divergence, denoted by $D_{KL}[. \mid\mid .]$ is 0. Hence, from

(7.13), it can be written that

$$-E_{Q(z_c,y,z)}\left[\log\left(\frac{P_{\mathcal{G}\mathcal{Z}_c\mathcal{Y}\mathcal{Z}}(x,z_c,y,z)}{Q(z_c,y,z)}\right)\right]+\log P(x)\geqslant 0$$

$$\Rightarrow \log P(x)\geqslant E_{Q(z_c,y,z)}\left[\log\left(\frac{P_{\mathcal{G}\mathcal{Z}_c\mathcal{Y}\mathcal{Z}}(x,z_c,y,z)}{Q(z_c,y,z)}\right)\right]. \tag{7.14}$$

Here, $\log P(x)$ denotes the log evidence and hence, the right hand side of inequality (7.14) represents the ELBO. Thus, it can be stated that minimizing the KL divergence in (7.13) is same as maximizing the ELBO in inequality (7.14).

To frame it as minimization objective, taking negative of inequality (7.14), the upper bound can be obtained as follows:

$$-\log P(x)\leqslant -E_{Q(z_c,y,z)}\left[\log\left(\frac{P_{\mathcal{G}\mathcal{Z}_c\mathcal{Y}\mathcal{Z}}(x,z_c,y,z)}{Q(z_c,y,z)}\right)\right]. \tag{7.15}$$

Let, the right hand side of inequality (7.15), which is to be minimized, be denoted by $\mathcal{M}$. Then,

$$\mathcal{M}=-E_{Q(z_c,y,z)}\left[\log\left(\frac{P_{\mathcal{G}\mathcal{Z}_c\mathcal{Y}\mathcal{Z}}(x,z_c,y,z)}{Q(z_c,y,z)}\right)\right].$$

Using (7.10), $\mathcal{M}$ can be rewritten as follows:

$$\mathcal{M}=-E_{Q(z_c,y,z)}\left[\log\left(\frac{P_{\mathcal{G}}(x\mid z_c,y,z)P_{\mathcal{Z}_c}(z_c)P_{\mathcal{Y}\mathcal{Z}}(y,z)}{Q(z_c,y,z_s)}\right)\right]$$

$$=-E_{Q(z_c,y,z)}[\log P_{\mathcal{G}}(x\mid z_c,y,z)]-E_{Q(z_c,y,z)}\left[\log\frac{P_{\mathcal{Z}_c}(z_c)}{Q(z_c,y,z)}\right]$$

$$-E_{Q(z_c,y,z)}\left[\log\frac{P_{\mathcal{Y}\mathcal{Z}}(y,z)}{Q(z_c,y,z)}\right]-E_{Q(z_c,y,z)}[\log Q(z_c,y,z)]$$

$$=-E_{Q(z_c,y,z)}[\log P_{\mathcal{G}}(x\mid z_c,y,z)]+D_{KL}[Q(z_c)\parallel P_{\mathcal{Z}_c}(z_c)]$$

$$+D_{KL}[Q(y,z)\parallel P_{\mathcal{Y}\mathcal{Z}}(y,z)]-E_{Q(z_c,y,z)}[\log Q(z_c,y,z)]. \tag{7.16}$$

where the last term in (7.16) represents the entropy defined over $Q(z_c,y,z)$, which acts as a regularizer for auxiliary distribution $Q(z_c,y,z)$. As z_c is assumed to be pairwise independent of both y and z, from optimization perspective, optimizing $D_{KL}[Q(z_c,y,z)\parallel P_{\mathcal{Z}_c}(z_c)]$ is equivalent to optimizing $D_{KL}[Q(z_c)\parallel P_{\mathcal{Z}_c}(z_c)]$. Similarly, $D_{KL}[Q(z_c,y,z)\parallel P_{\mathcal{Y}\mathcal{Z}}(y,z)]$ can be optimized by optimizing $D_{KL}[Q(y,z)\parallel P_{\mathcal{Y}\mathcal{Z}}(y,z)]$. Thus, from 7.16, the reconstruction objective to be minimized can be framed as:

$$\mathcal{J}_{\text{rec}}=\underbrace{-E_{Q(z_c,y,z)}[\log P_{\mathcal{G}}(x\mid z_c,y,z)]}_{L_R}-E_{Q(z_c,y,z)}[\log Q(z_c,y,z)]$$

$$+ \underbrace{D_{KL}[Q(z_c) \parallel P_{\mathcal{Z}_c}(z_c)]}_{R_1} + \underbrace{D_{KL}[Q(y, z) \parallel P_{\mathcal{Y}\mathcal{Z}}(y, z)]}_{R_2}, \tag{7.17}$$

where L_R represents the reconstruction loss, whereas R_1 and R_2 denote the regularization terms corresponding to color appearance code z_c, and joint density of nuclei segmentation map y and embedding information z, respectively.

To solve the minimization problem in (7.17), the priors for latent distributions, corresponding to color appearance code z_c as well as joint density of nuclei segmentation map y and embedding information z, are to be assumed in order to compute the KL divergence values. Now, the latent color appearance representation, corresponding to each stained histological image, should capture information corresponding to all the stains involved in the staining routine. Hence, similar to Chapter 5, a mixture of truncated normal distributions, is used as the prior for latent distribution corresponding to color appearance code. As nuclei segmentation map and embedding information, corresponding to a particular histological image, is to be kept intact after color normalization, standard normal distribution is used as the prior for the joint density of segmentation map and embedding information. To ensure that the histological information is contained after reconstruction, a structural similarity based loss term $l_{SSIM}(\mathcal{G}, \mathcal{E}_c, \mathcal{F}_\phi, \mathcal{E}_\omega)$ is incorporated along with the reconstruction loss $\mathcal{J}_{\text{rec}}$. The loss term l_{SSIM} is defined as follows:

$$l_{SSIM}(\mathcal{G}, \mathcal{E}_c, \mathcal{F}_\phi, \mathcal{E}_\omega) = 1 - SSIM[x, \mathcal{G}(\mathcal{E}_c(x), \mathcal{F}_\phi(x), \mathcal{E}_\omega(x))] \tag{7.18}$$

So, the overall reconstruction loss to be minimized is given as follows:

$$J_{\text{rec}} = \mathcal{J}_{\text{rec}} + l_{SSIM}(\mathcal{G}, \mathcal{E}_c, \mathcal{F}_\phi, \mathcal{E}_\omega) \tag{7.19}$$

where SSIM denotes a structural similarity measurement index, called structural similarity index measure [129]. The simultaneous nuclei segmentation and color normalization algorithm is presented in Algorithm 7.1.

7.3 Performance Analysis

This section presents the performance of the simultaneous nuclei segmentation and color normalization model, termed as SNSCN in this chapter. The performance of the SNSCN model is compared with that of

- several state-of-the-art methods for stain estimation and stain separation: plane fitting (PF) [84], HTN due to Li and Plataniotis [132], enhanced PF (EPF) [89], structure-preserving color normalization (SPCN) [22], expectation-maximization (EM) algorithm [134], and rough-fuzzy circular clustering method RFCC$_{vM}$ [105];

- several state-of-the-art color normalization methods, such as, color transfer (ColTrans) [38], stain color description (SCD) [15], SN-GAN model [41], StainGAN [90], AST model [6], as well as PF [84], HTN [132], EPF [89], SPCN [22] and TredMiL [118]; and

---

Algorithm 7.1 Simultaneous nuclei segmentation and color normalization of non-normalized source histological images.

Input: A set of N non-normalized source images $\{x_n^S\}_{n=1}^N$, template image x^T, trained network parameters $\{\theta_{\mathcal{E}_c}, \theta_{\mathcal{F}_\phi}, \theta_{\mathcal{E}_\omega}, \theta_{\mathcal{G}}\}$ corresponding to deep networks $\mathcal{E}_c$, $\mathcal{F}_\phi$, $\mathcal{E}_\omega$ and $\mathcal{G}$.

Output: Color normalized source images $\{\tilde{x}_n^S\}_{n=1}^N$ and nuclei segmentation map $\{\tilde{y}_n^S\}_{n=1}^N$.

1: Generate latent color appearance code z_c^T, and nuclei segmentation map y^T and embedding information z^T corresponding to template image x^T by using $\mathcal{E}_c(x^T; \theta_{\mathcal{E}_c})$, $\mathcal{F}_\phi(x^T; \theta_{\mathcal{F}_\phi})$ and $\mathcal{E}_\omega(x^T; \theta_{\mathcal{E}_\omega})$, respectively.

2: **for** each image in source image set $\{x_n^S\}_{n=1}^N$ **do**

  • Generate latent color appearance code $z_{c_n}^S$, nuclei segmentation map y_n^S and embedding information z_n^S by using $\mathcal{E}_c(x_n^S; \theta_{\mathcal{E}_c})$, $\mathcal{F}_\phi(x_n^S; \theta_{\mathcal{F}_\phi})$ and $\mathcal{E}_\omega(x_n^S; \theta_{\mathcal{E}_\omega})$, respectively.

  • Feed template image latent color appearance code z_c^T, and source image nuclei segmentation map y^S and embedding information z^S to the decoder $\mathcal{G}$ and by using $\mathcal{G}(z_c^T, y^S, z^S; \theta_{\mathcal{G}})$ generate normalized source image $\tilde{x}_n^S$.

  • Feed the normalized source image $\tilde{x}_n^S$ to the nuclei segmentation map generator $\mathcal{F}_\phi$ to generate final nuclei segmentation map $\tilde{y}_n^S$.

3: **end for**

4: Stop.

  • several existing nuclei segmentation models: U-Net [40], Mask-R-CNN [66], U-Net++ [150]; and state-of-the-art nuclei segmentation methods, namely, HoVer-Net [114], multi-organ nuclei segmentation (MoNS) [43], Stardist [93], WNSeg [79], Swin-MIL [148], BoNuS [140] and Ostrich model proposed in Chapter 6.

To analyze the performance of the proposed SNSCN model and existing models in stain color normalization, publicly available UCSB data set [36] is used in this study. A brief description of the UCSB data set is provided in Chapter 3 . The performance of the SNSCN model and several state-of-the-art algorithms in nuclei segmentation is analyzed using the publicly available TCGA data set [97]. Brief description of the TCGA data set is provided in Chapter 6. The proposed model is trained with $2,535$ overlapping image patches of size 256×256 corresponding to 15 training set images of TCGA data set, while the validation set contains 507 overlapping patches and the test set comprises of 192 non-overlapping image patches. The SNSCN model is trained using an NVIDIA RTX A4000 GPU with 6144 CUDA cores and 16 GB storage. The total time taken to train the proposed SNSCN model on $2,535$ training set images of TCGA data set is around 38 hours approximately.

To analyze the performance of different variants of the SNSCN model and several state-of-the-art methods in stain estimation, element-wise standard deviation is considered similar to Chapter 3, Chapter 4 and Chapter 5. On the other hand, the performance of different color normalization methods is evaluated using normalized median intensity (NMI) [74], between-image color constancy (BiCC) index and within-set color constancy (WsCC) index [105]. Similar to Chapter 6, four evaluation indices, namely, Dice coefficient, Jaccard score, precision and recall, have been used to evaluate the performance of nuclei segmentation.

7.3.1 Performance Analysis in Color Normalization

The effectiveness of the SNSCN model in color normalization is established through the following analyses.

(a) Hematoxylin

(b) Eosin
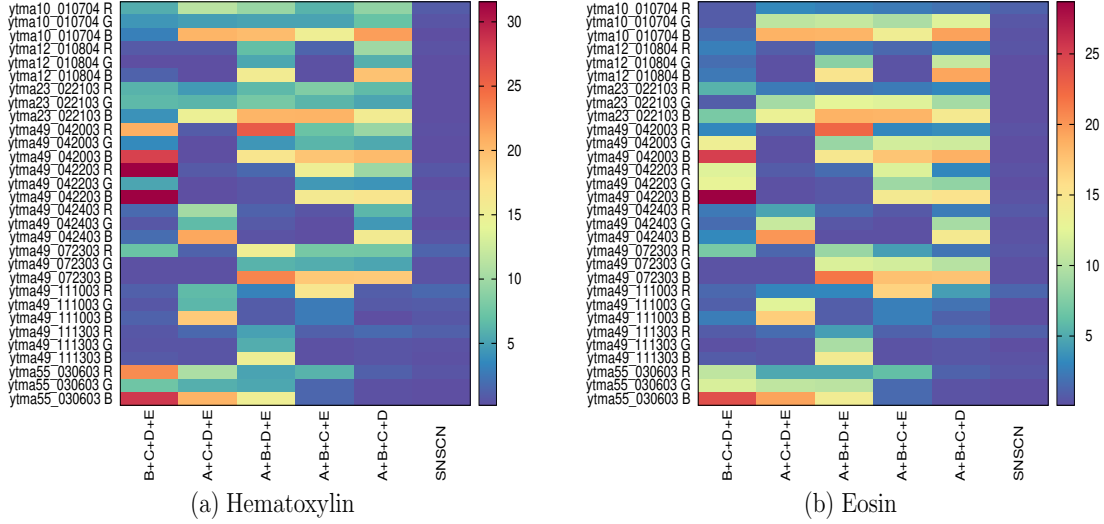
Figure 7.2: Performance analysis of stain vector estimation of the proposed SNSCN model and all five combinations of constituent terms of objective function.
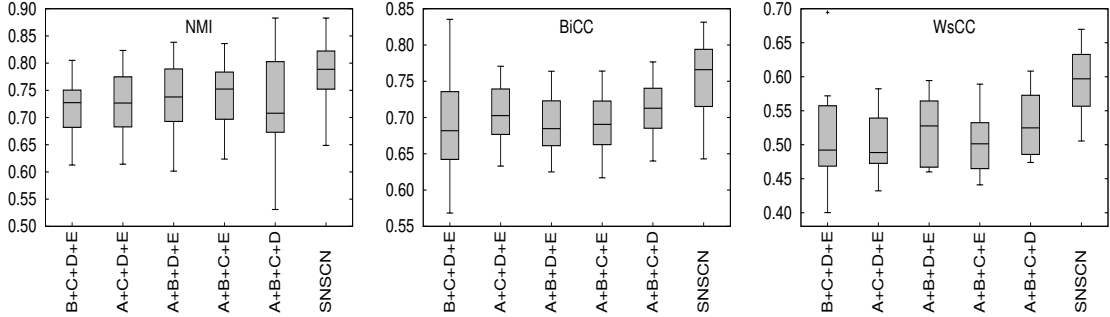
Figure 7.3: Performance analysis of color normalization of the proposed SNSCN model and all five combinations of constituent terms of objective function.

7.3.1.1 Ablation Study

Apart from the adversarial term J_{adv}, the objective function of the SNSCN model contains five terms. Let, these five terms, attributed by the reconstructive module J_{rec}, be denoted as: $A = -E_{Q(z_c,y,z)}[\log P_{\mathcal{G}}(x \mid z_c, y, z)]$, $B = D_{KL}[Q(z_c) \parallel P_{\mathcal{Z}_c}(z_c)]$, $C = D_{KL}[Q(y,z) \parallel P_{\mathcal{Y}\mathcal{Z}}(y,z)]$, $D = -E_{Q(z_c,y,z)}[\log Q(z_c,y,z)]$ and $E = l_{SSIM}(\mathcal{G}, \mathcal{E}_c, \mathcal{F}_\phi, \mathcal{E}_\omega)$. The significance of each individual term is established through an ablation study, where each term is omitted from the objective function and the performance of the model in stain vector estimation is observed in the absence of that particular term. From the heatmap representation presented

Table 7.1: Statistical Significance Analysis of Different Algorithms

Different	NMI		BiCC		WsCC	
Methods	Wilcoxon	Paired-t	Wilcoxon	Paired-t	Wilcoxon	Paired-t
B+C+D+E	2.57E-11	4.98E-17	1.58E-07	4.44E-09	4.67E-03	9.19E-04
A+C+D+E	1.75E-11	2.40E-14	4.24E-11	7.55E-16	2.53E-03	4.34E-07
A+B+D+E	1.20E-09	5.63E-11	1.85E-11	3.52E-19	2.53E-03	7.52E-06
A+B+C+E	3.77E-11	2.28E-10	1.85E-11	2.31E-20	2.53E-03	5.85E-06
A+B+C+D	2.63E-08	2.91E-10	2.40E-11	6.78E-19	2.53E-03	1.01E-07
Correlated	4.02E-11	4.07E-12	1.75E-11	9.88E-19	2.53E-03	1.47E-05
ColTrans	1.75E-11	9.23E-23	1.75E-11	8.62E-30	2.53E-03	9.81E-10
PF	2.53E-11	3.57E-21	1.75E-11	1.50E-24	2.53E-03	4.87E-08
EPF	1.85E-11	7.73E-23	1.75E-11	8.49E-25	2.53E-03	1.92E-08
SCD	7.43E-11	2.34E-16	1.95E-11	2.32E-23	2.53E-03	4.39E-05
HTN	7.76E-09	2.81E-11	1.17E-10	1.23E-15	3.46E-03	2.38E-04
SPCN	1.75E-11	1.80E-23	1.75E-11	1.45E-26	2.53E-03	9.45E-08
SN-GAN	5.67E-10	7.60E-15	1.85E-11	7.89E-23	2.53E-03	2.03E-06
StainGAN	1.75E-11	5.60E-25	1.75E-11	1.98E-30	2.53E-03	6.90E-09
AST	1.85E-11	2.36E-19	1.75E-11	3.51E-25	2.53E-03	7.96E-07
TredMiL	2.69E-07	3.16E-06	4.30E-09	4.26E-11	2.53E-03	1.71E-04

in Fig. 7.2(a), it is evident that the proposed SNSCN model with all five constituent terms performs better than every individual combination in the estimation of H-stain vector. Similarly, the heatmap presented in Fig. 7.2(b) ensures that the SNSCN model outperforms every individual combination in case of E-stain vector estimation.

The performance of the proposed SNSCN model in color normalization is also compared with that of other five combinations of the constituent terms of the objective function with respect to NMI, BiCC and WsCC, and the corresponding results are reported in Fig. 7.3. From the boxplot representation presented in Fig. 7.3, it can be observed that SNSCN attains highest median with respect to all three quantitative indices used. The statistical significance of the SNSCN model is analyzed with respect to p-values computed through Wilcoxon signed-rank (one-tailed) and paired-t (one-tailed) tests. From the p-values reported in Table 7.1, it can be observed that the SNSCN model performs significantly better than every individual constituent terms of the objective function, considering 95% confidence level.

7.3.1.2 Pairwise Independence Between z_c, and y and z

The SNSCN model assumes that the latent color appearance information z_c is pairwise independent of nuclei segmentation map y and embedding information z. The performance of the proposed model in stain vector estimation is compared with that of the model where z_c, y and z are correlated, and the corresponding results are reported in Fig. 7.4. It can be observed from the heatmap based comparison presented in Fig. 5.4(a) that corresponding to H-stain, the proposed model outperforms the model where z_c, y and z are correlated. Similarly, from Fig. 7.4(b), it can also be noticed that in case of E-stain also, SNSCN performs better than the model with correlated z_c, and y and z.

To evaluate the efficacy of the proposed model in color normalization, its performance is compared with that of the model where z_c, y and z are correlated. From the boxplot

(a) Hematoxylin   (b) Eosin
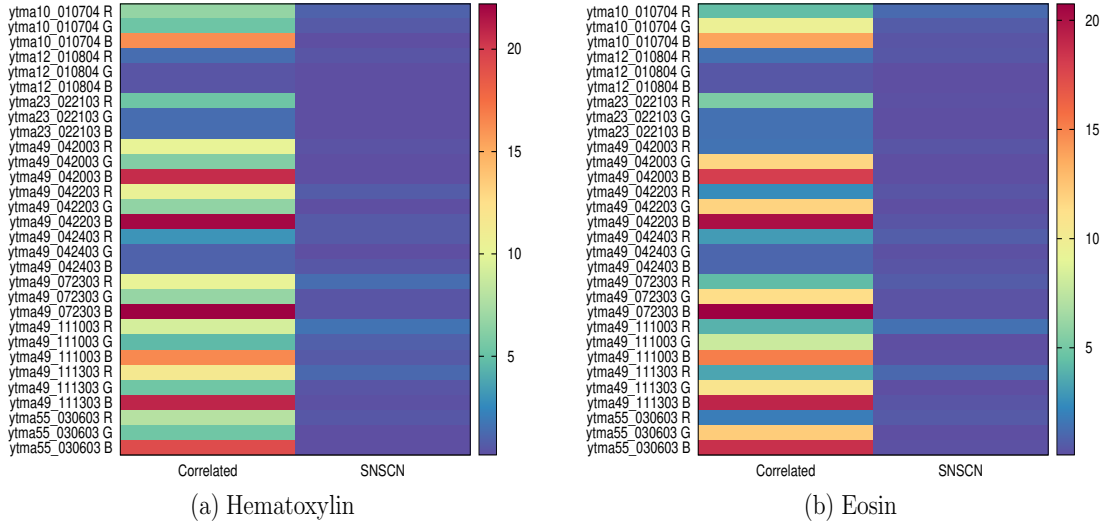
Figure 7.4: Performance analysis of stain vector estimation of proposed SNSCN model and the variant where color appearance code z_c, and nuclei segmentation map y and embedding information z are correlated.
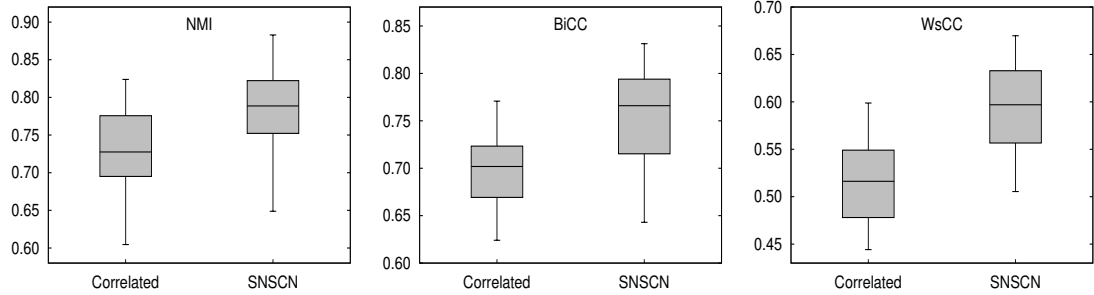
Figure 7.5: Performance analysis of color normalization of the proposed SNSCN model and the variant where color appearance code z_c, and nuclei segmentation map y and embedding information z are correlated.

representation provided in Fig. 7.5, it is evident that the SNSCN model with independent z_c, and y and z attains higher median with respect to NMI, BiCC and WsCC than the model where z_c, y and z are assumed to be correlated. As the model with correlated z_c, y and z loses significant amount of structural information during mapping, the proposed model with independent z_c, y and z outperforms the correlated model. From the p-values reported in Table 7.1, it is evident that the proposed SNSCN model performs significantly better than the model where z_c, y and z are correlated.

7.3.1.3   Comparison with State-of-the-Art Methods

The performance of the proposed model in stain vector estimation is compared with that of different state-of-the-art methods, and the results, corresponding to H-stain and E-stain, are reported via heatmap representations presented in Fig. 7.6(a) and Fig. 7.6(b),

(a) Hematoxylin	(b) Eosin
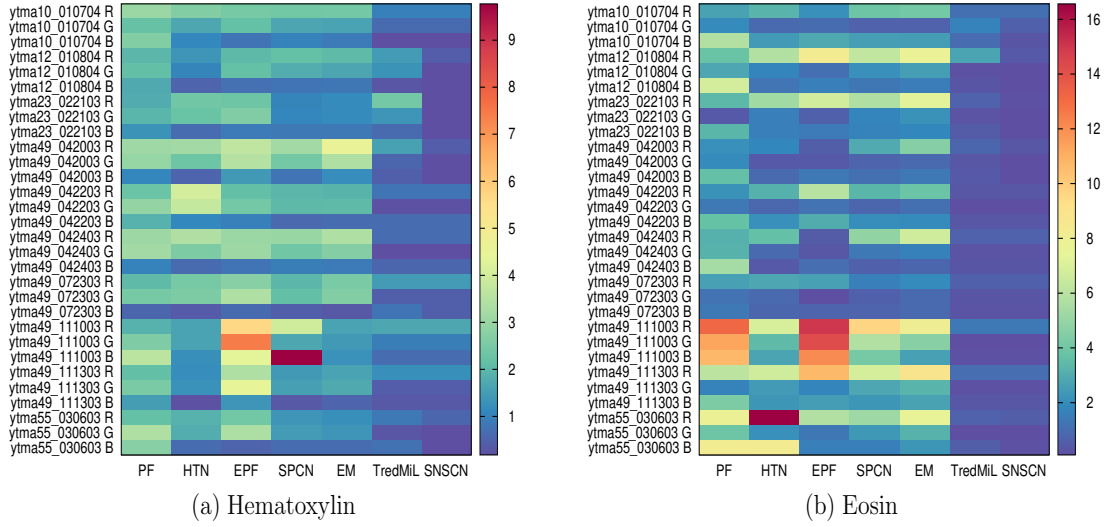
Figure 7.6: Comparative performance analysis of stain vector estimation of the SNSCN model and (a)-(b) different existing methods such as PF, HTN, EPF, SPCN, EM and TredMiL in stain vector estimation (H-stain and E-stain, respectively).
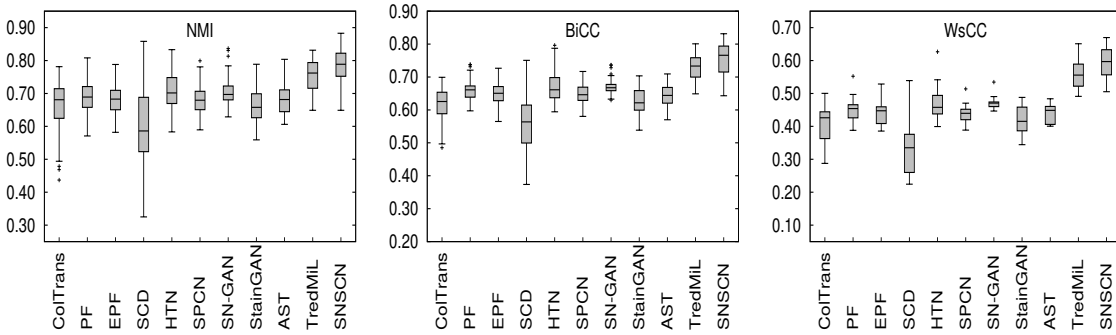
Figure 7.7: Comparative performance analysis of color normalization of the SNSCN model and state-of-the-art methods such as ColTrans, PF, EPF, SCD, HTN, SPCN, SN-GAN, StainGAN, AST and TredMiL in color normalization.

respectively. From the heatmap presented in Fig. 7.6(a), it can be observed that the SNSCN model performs better than state-of-the-art stain vector estimation methods in case of H-stain. From Fig. 7.6(b), it is also evident that SNSCN achieves lowest σ values in most of the cases compared to the state-of-the-art methods in the estimation of E-stain representative vector.

The results presented in Fig. 7.7 ensure the fact that the proposed model exhibits highest median values than all the existing methods with respect to all three quantitative indices, namely, NMI, BiCC and WsCC. Analyzing the p-values reported in Table 7.1, it can be also be observed that the proposed method performs significantly better than the state-of-the-art color normalization methods. The qualitative performance of different color normalization algorithms is presented in Fig. 7.8. The results reported in Table 7.1, Fig. 7.7 and Fig. 7.8 establish the fact that the SNSCN model outperforms state-of-the-art

136
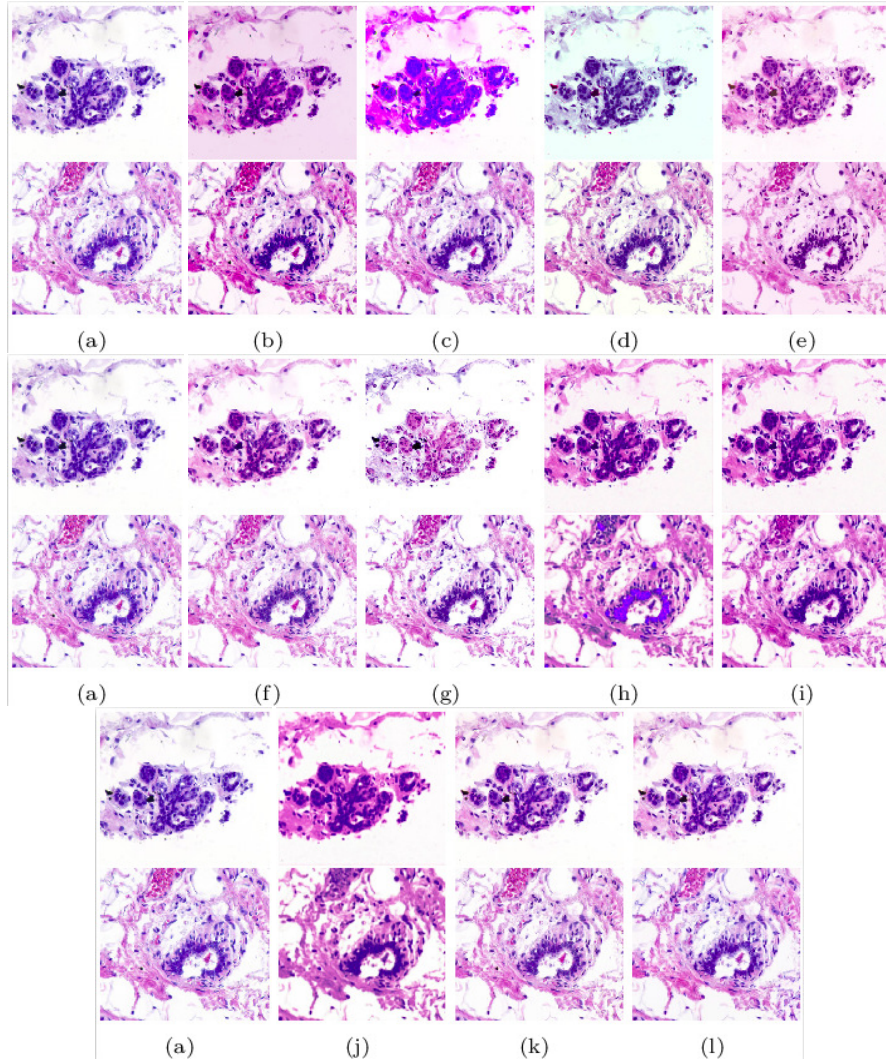
Figure 7.8: (a) Original images of UCSB data set; and color normalized images obtained using different state-of-the-art color normalization methods: (b) ColTrans, (c) PF, (d) EPF, (e) SCD, (f) HTN, (g) SPCN, (h) SN-GAN, (i) StainGAN, (j) AST, (k) TredMiL and (l) SNSCN.

color normalization algorithms as per as color consistency after normalization is concerned.

7.3.2 Performance Analysis in Nuclei Segmentation

The performance of the SNSCN model is finally compared with that of several state-of-the-art nuclei segmentation approaches on TCGA data set and the results with respect to different quantitative indices are reported in Table 7.2. From the results reported in Table 7.2, it can be observed that the proposed model outperforms all the state-of-the-art nuclei segmentation approaches, with respect to all the evaluation indices, namely, Dice, Jaccard, precision and recall. The qualitative performance of the SNSCN model, along

Table 7.2: Quantitative performance analysis in nuclei segmentation on TCGA Data set: State-of-the-art models vs SNSCN

Methods	Dice	Jaccard	Precision	Recall
SNSCN	**0.788345**	**0.653957**	**0.826117**	**0.786249**
U-Net	0.645650	0.446128	0.624558	0.668216
Mask-R-CNN	0.747086	0.585659	0.799975	0.700757
U-Net++	0.773155	0.620294	0.796708	0.750955
HoVer-Net	0.744516	0.586780	0.811804	0.687529
MoNS	0.755509	0.614706	0.748012	0.763157
Stardist	0.743228	0.585713	0.811579	0.685495
WNSeg	0.773406	0.625322	0.813470	0.737104
Swin-MIL	0.749044	0.593266	0.801989	0.702656
BoNuS	0.784740	0.639496	0.806351	0.764257
Ostrich	0.788035	0.641520	0.813523	0.764096

Table 7.3: Statistical significance analysis of different algorithms with respect to Paired-t (one-tailed) test on TCGA data

Methods	Dice	Jaccard	Precision	Recall
U-Net	3.89E-47	4.37E-49	1.14E-15	3.77E-51
Mask-R-CNN	3.44E-31	1.01E-37	8.13E-81	1.83E-08
U-Net++	8.79E-40	7.05E-41	4.70E-25	4.24E-26
HoVer-Net	2.41E-52	5.58E-57	2.52E-12	3.10E-67
MoNS	8.16E-28	2.12E-29	1.26E-60	1.93E-01
Stardist	5.45E-34	1.63E-36	1.15E-04	7.20E-42
WNSeg	1.16E-48	1.17E-49	4.20E-16	4.21E-63
Swin-MIL	4.42E-51	2.95E-57	1.29E-16	5.34E-67
BoNuS	3.84E-17	4.03E-18	9.68E-28	7.56E-26
Ostrich	7.54E-20	1.46E-19	1.95E-24	4.12E-28

Table 7.4: Statistical significance analysis of different algorithms with respect to Wilcoxon signed-rank (one-tailed) test on TCGA data

Methods	Dice	Jaccard	Precision	Recall
U-Net	7.85E-32	8.61E-32	2.38E-14	6.25E-32
Mask-R-CNN	1.30E-32	1.19E-32	1.86E-33	1.71E-02
U-Net++	2.70E-29	2.44E-29	1.68E-25	1.42E-21
HoVer-Net	3.22E-33	3.22E-33	5.00E-11	1.49E-33
MoNS	7.62E-24	3.46E-24	1.47E-33	1.36E-01
Stardist	7.12E-28	6.64E-28	1.14E-05	4.17E-30
WNSeg	6.44E-32	6.34E-32	7.51E-15	1.62E-33
Swin-MIL	1.92E-33	1.89E-33	1.84E-18	1.47E-33
BoNuS	4.16E-16	1.98E-16	4.03E-24	7.74E-23
Ostrich	3.07E-21	3.07E-21	1.48E-22	3.36E-24

with a comparison with state-of-the-art approaches, in nuclei segmentation, is presented in Fig. 7.9 using image patches from the TCGA data set. Analyzing the p-values reported in Table 7.3, it can be observed that, with respect paired-t test, the proposed SNSCN model performs significantly better in all the cases. The p-values reported in Table 7.4 also depict
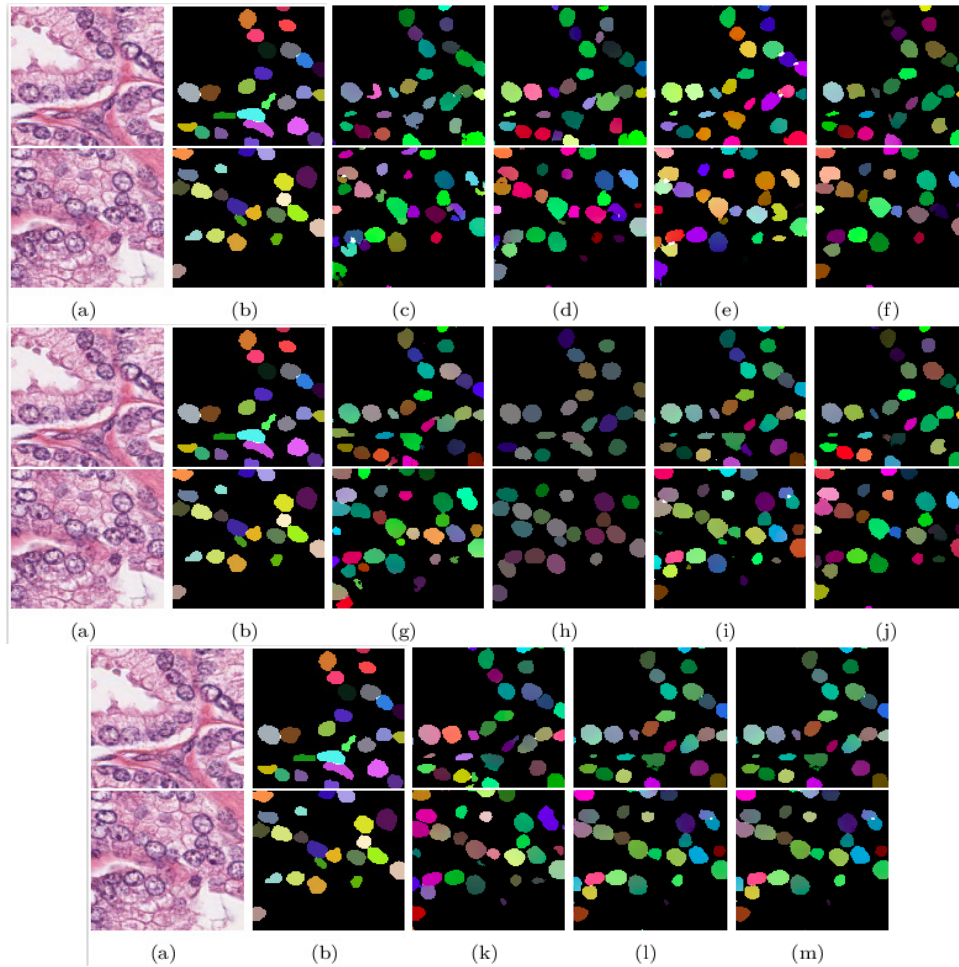
Figure 7.9: (a) Original images of TCGA data set, (b) ground-truth segmentation map; and segmentation maps obtained using existing nuclei segmentation methods: (c) U-Net, (d) Mask-R-CNN, (e) U-Net++, (f) HoVer-Net, (g) MoNS, (h) Stardist, (i) WNSeg, (j) Swin-MIL, (k) BoNuS, (l) Ostrich and (m) SNSCN.

the fact that the proposed model performs significantly better in all the cases, with respect Wilcoxon signed-rank test.

7.4   Conclusion

The problem of nuclei segmentation from histological images, in presence of inadmissible disagreement in the appearance of stained tissue color, is of high importance as the variation in color appearance degrades the performance of nuclei segmentation algorithm. In this regard, the major contribution of this chapter is to introduce a novel algorithm, called SNSCN, for simultaneous nuclei segmentation and color normalization of histological images. The proposed model deals with the overlapping nature of histochemical stains, by incorporating a mixture of truncated normal distributions as the prior for latent color ap-

pearance code. From the quantitative as well as qualitative results reported in this chapter, it can be observed that the proposed model outperforms state-of-the-art nuclei segmentation and stain color normalization algorithms. The proposed method performs better than other existing color normalization methods, in terms of handling overlapping information, and ensuring within- image and between-image color constancy of nuclei structures after color normalization. The results reported in this chapter also validates the fact that nuclei segmentation on stain color normalized histological images improves the segmentation accuracy.

Chapter 8

Conclusion and Future Directions

This chapter provides a brief summary of the main contributions of the research work, presented in different chapters of the thesis. The future research directions, along with possible extensions and applications, are also discussed in this chapter.

8.1 Major Contributions

The thesis focuses on developing some novel models to deal with the nuclei segmentation and color normalization problem of histological images. The significant challenges associated with nuclei segmentation and color normalization of histological images include (i) separation of staining information from overlapped appearance of histochemical stains, (ii) extraction of color appearance information independent of stain density information such that the loss or modification in color appearance does not affect stain density information, and (iii) devising a nuclei segmentation algorithm, generalizable to different types of image representations in segmentation map domains. All the aforementioned issues have been addressed in this thesis. A concise summary, emphasizing on the important features of the proposed models, is presented next.

Chapter 3 introduces a novel circular clustering algorithm, termed as rough-fuzzy circular clustering ($RFCC_{cosine}$), which assumes that each stain class is represented by a centroid, a crisp lower approximation, and a fuzzy boundary. The proposed $RFCC_{cosine}$ method utilizes saturation-weighted hue histogram instead of standard hue histogram to reduce the impact of less-saturated achromatic pixels in stain color estimation. From the quantitative and qualitative results reported in Chapter 3, it can be noted that, rough-fuzzy circular clustering, being a generalization of existing algorithms, performs well in reducing the variation in the color appearance of histological images. However, the cosine distance, used in $RFCC_{cosine}$ method, cannot model concentration of hue values near the mean of each stain class and eventually fails to capture the underlying data distribution properly. In this regard, a new dissimilarity measure is introduced in Chapter 4 to deal with the circular nature of the hue values. Based on the new dissimilarity measure, a new rough-fuzzy circular clustering algorithm, termed as $RFCC_{vM}$, is devised in this chapter to address the disagreement problem in the appearance of stained tissue color. It is evident from the quantitative and qualitative results presented in Chapter 4 that the integration

of rough-fuzzy circular clustering with von Mises distribution provides better modeling of circular data and enables proper analysis of stained histological images.

To ensure the element-wise non-negativity of the estimated stain color appearance matrix and associated stain density map, both $RFCC_{cosine}$ and $RFCC_{vM}$ rely on non-negative matrix factorization method, which itself suffers from unstable convergence problem. Moreover, the correlation between color appearance information and stain density information is not explored properly in Chapter 3 and Chapter 4. In this respect, to normalize color appearance without hampering the histological information contained in an image, a new deep generative model, termed as TredMiL, has been introduced in Chapter 5. To deal with the overlapped appearance of histochemical stains, TredMiL model considers a mixture of truncated normal distribution as the prior for latent color appearance code. The results reported in Chapter 5 highlight the fact that TredMiL performs consistently well in the cases where color appearance information of an image is somehow corrupted or lost. This is due to the fact that TredMiL assumes the color appearance information and stain bound information to be independent of each other. An important result is that the trained TredMiL model can be applied to any data set with H&E staining, which is one of the most commonly used staining methods in histopathological image analysis.

An important result presented in Chapter 5 demonstrates that color normalization has a huge impact on subsequent nuclei segmentation task. Mitigating the disagreement in the appearance of stained tissue color improves quality of region-of-interest (ROI) segmentation from histological images. After addressing the color normalization problem in Chapter 5, a novel deep generative model, termed as Ostrich, is introduced in Chapter 6. The Ostrich model considers an embedding space to balance information disparity between information-rich histological image space and information-poor segmentation map domain. From the quantitative and qualitative results reported in Chapter 6, it is evident that incorporation of embedding space balances the information gap between two image domains and enables the model to perform better than other image-to-image translation networks. The theoretical analysis using optimal transport and measure theory for reducing the network complexity is also validated by the experimental results. An important aspect is that image-to-image translation paradigm helps Ostrich to perform consistently well in different nuclei segmentation tasks with different types of target domain image representations.

It has been shown in Chapter 5 that stain color normalization enables better nuclei segmentation from histological images. On the other hand, accurate nuclei segmentation makes color normalization rather trivial. In this regard, Chapter 7 presents a novel model, termed as SNSCN, for simultaneous nuclei segmentation and color normalization of histological images by integrating the concept of spatial attention and the merit of truncated normal distribution. Being inspired by the success of TredMiL model, introduced in Chapter 5, in color normalization, the SNSCN model assumes that the latent appearance code is sampled from a mixture of truncated normal distributions. The comparative performance analysis presented in Chapter 7 demonstrates that the proposed SNSCN model performs better in nuclei segmentation as well as color normalization, compared to the existing nuclei segmentation and color normalization methods.

8.2 Future Directions

There are several important characteristics of the research work presented in this thesis that can be extended for the advancement of computer-aided histological image analysis. Some improvements and future directions are enlisted as follows:

1. The effectiveness of the proposed color normalization and nuclei segmentation frameworks is studied extensively on light absorber histochemical reagents, such as, hematoxylin and eosin stains. However, a thorough understanding of immunohistochemical stained histological slides is required before applying the proposed methods to H&DAB or other types of staining, where the stains are not light absorbers, rather they scatter light.

2. Another important future research direction can be to study the effect of color normalization in other types of downstream tasks in digital pathology such as cancerous cell classification or clinical end-point prediction.

3. Explainable AI (XAI) models can be developed for nuclei segmentation in order to significantly increase the transparency and reliability in computer-aided diagnostics. By studying feature importance and decision pathways, the pathologists can better comprehend and evaluate the conclusions made by AI models, which can boost AI's acceptance in clinical settings.

4. The histological images can be integrated to other imaging modalities, such as, MRI or CT. The multi-modal integration of imaging as well as non-imaging modalities may improve segmentation performance and provide comprehensive understanding of the pathological conditions.

5. Generative adversarial networks, used in color normalization and nuclei segmentation models proposed in this thesis, can be further utilized for realistic synthetic data generation, which can augment histological image data sets, mitigating the data-scarcity issues.

6. Active learning and crowdsourcing can improve the amount and quantity of annotated data sets. The manual workload of the pathologists can be reduced by utilizing automated annotation methods.

7. Finally, the segmentation frameworks can be further extended to deal with the segmentation and delineation of tissue components other than nuclei structures.

List of Related Publications

International Journal Papers

J1. **Suman Mahapatra** and Pradipta Maji. Optimal Transport Driven Asymmetric Image-to-Image Translation for Nuclei Segmentation of Histological Images. ***IEEE Transactions on Artificial Intelligence (under review)***, pages 1-12. Manuscript ID: TAI-2024-Nov-A-01785.

J2. **Suman Mahapatra** and Pradipta Maji. Truncated Normal Mixture Prior Based Deep Latent Model for Color Normalization of Histology Images. ***IEEE Transactions on Medical Imaging***, 42(6), pages 1746-1757, 2023. DOI:10.1109/TMI.2023.3238425.

J3. Pradipta Maji and **Suman Mahapatra**. Circular Clustering in Fuzzy Approximation Spaces for Color Normalization of Histological Images. ***IEEE Transactions on Medical Imaging***, 39(5), pages 1735-1745, 2020. DOI:10.1109/TMI.2019.2956944.

J4. Pradipta Maji and **Suman Mahapatra**. Rough-Fuzzy Circular Clustering for Color Normalization of Histological Images. ***Fundamenta Informaticae***, 164(1), pages 103-117, 2019. DOI:10.3233/FI-2019-1756.

References

[1] https://www.who.int/.

[2] G. James. *Modern Engineering Mathematics*. Pearson Education Limited, 2015.

[3] A. Arbelle and T. R. Raviv. Microscopy Cell Segmentation via Adversarial Neural Networks. In *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*, pages 645–648. IEEE, 2018.

[4] A. Basavanhally and A. Madabhushi. EM-Based Segmentation-Driven Color Standardization of Digitized Histopathology. In *Proceedings of SPIE Medical Imaging*, volume 8676, pages 86760G–1–86760G–12, 2013.

[5] A. Basavanhally, M. Feldman, N. Shih, C. Mies, J. Tomaszewski, S. Ganesan and A. Madabhushi. Multi-field-of-view Strategy for Image-based Outcome Prediction of Multi-parametric Estrogen Receptor-positive Breast Cancer Histopathology: Comparison to Oncotype DX. *Journal of pathology informatics*, 2(2):1, 2012.

[6] A. Bentaieb and G. Hamarneh. Adversarial Stain Transfer for Histopathology Image Analysis. *IEEE Transactions on Medical Imaging*, 37(3):792–802, 2018.

[7] A. C. Ruifrok, D. A. Johnston and others. Quantification of Histochemical Staining by Color Deconvolution. *Analytical and Quantitative Cytology and Histology*, 23(4):291–299, 2001.

[8] A. C. Ruifrok, R. L. Katz and D. A. Johnston. Comparison of Quantification of Histochemical Staining by Hue-Saturation-Intensity (HSI) Transformation and Color-Deconvolution. *Applied Immunohistochemistry & Molecular Morphology*, 11(1):85–91, 2003.

[9] A. Gupta, R. Duggal, S. Gehlot, R. Gupta, A. Mangal, L. Kumar, N. Thakkar and D. Satpathy. GCTI-SN: Geometry-inspired Chemical and Tissue Invariant Stain Normalization of Microscopic Medical Images. *Medical Image Analysis*, 65:101788, 2020.

[10] A. Hanbury. Circular Statistics Applied to Colour Images. In *Proceedings of 8th Computer Vision Winter Workshop*, volume 91, pages 53–71, 2003.

[11] A. Janowczyk, A. Basavanhally and A. Madabhushi. Stain Normalization Using Sparse Autoencoders (StaNoSA): Application to Digital Pathology. *Computerized Medical Imaging and Graphics*, 57:50–61, 2017.

[12] A. Janowczyk, S. Doyle, H. Gilmore and A. Madabhushi. A resolution Adaptive Deep Hierarchical (RADHicaL) Learning Scheme Applied to Nuclear Segmentation of Digital Pathology Images. *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization*, 6(3):270–276, 2018.

[13] A. Kendall, V. Badrinarayanan and R. Cipolla. Bayesian Segnet: Model Uncertainty in Deep Convolutional Encoder-Decoder Architectures for Scene Understanding. *arXiv preprint arXiv:1511.02680*, 2015.

[14] A. Krizhevsky, I. Sutskever and G. E. Hinton. Imagenet Classification with Deep Convolutional Neural Networks. *Advances in Neural Information Processing Systems*, 25, 2012.

[15] A. M. Khan, N. Rajpoot, D. Treanor and D. Magee. A Nonlinear Mapping Approach to Stain Normalization in Digital Histopathology Images Using Image-Specific Color Deconvolution. *IEEE Transactions on Biomedical Engineering*, 61(6):1729–1738, 2014.

[16] A. Mahbod et al. Improving Generalization Capability of Deep Learning-based Nuclei Instance Segmentation by Non-deterministic Train Time and Deterministic Test Time Stain Normalization. *Computational and Structural Biotechnology Journal*, 23:669–678, 2024.

[17] A. Mahbod, G. Schaefer, I. Ellinger, R. Ecker, Ö. Smedby and C. Wang. A Two-Stage U-Net Algorithm for Segmentation of Nuclei in H&E-Stained Tissues. In *Proceedings of Digital Pathology: 15th European Congress, ECDP 2019, Warwick, UK*, pages 75–82. Springer, 2019.

[18] A. Patil et al. Fast, Self Supervised, Fully Convolutional Color Normalization of H&E Stained Images. In *Proceedings of IEEE International Symposium on Biomedical Imaging (ISBI)*, pages 1563–1567, 2021.

[19] A. Rabinovich, S. Agarwal, C. Laris, J. H. Price and S. J. Belongie. Unsupervised Color Decomposition of Histologically Stained Tissue Samples. In *Proceedings of Advances in Neural Information Processing Systems*, pages 667–674. MIT Press, 2004.

[20] A. Tabesh, M. Teverovskiy, H. Y. Pang, V. P. Kumar, D. Verbel, A. Kotsianti and O. Saidi. Multifeature Prostate Cancer Diagnosis and Gleason Grading of Histological Images. *IEEE Transactions on Medical Imaging*, 26(10):1366–1378, 2007.

[21] A. Tareef et al. Optimizing The Cervix Cytological Examination Based on Deep Learning and Dynamic Shape Modeling. *Neurocomputing*, 248:28–40, 2017.

[22] A. Vahadane, T. Peng, A. Sethi, S. Albarqouni, L. Wang, M. Baust, K. Steiger, A. M. Schlitter, I. Esposito and N. Navab. Structure-Preserving Color Normalization and Sparse Stain Separation for Histological Images. *IEEE Transactions on Medical Imaging*, 35(8):1962–1971, 2016.

[23] A. Vahadane, T. Peng, S. Albarqouni, M. Baust, K. Steiger, A. M. Schlitter, A. Sethi, I. Esposito and N. Navab. Structure-Preserved Color Normalization for Histological

Images. In *Proceedings of IEEE 12th International Symposium on Biomedical Imaging (ISBI)*, pages 1012–1015, 2015.

[24] B. E. Bejnordi, G. Litjens, N. Timofeeva, I. Otte-Hã́ller, A. Homeyer, N. Karssemeijer and J. A. W. M. van der Laak. Stain Specific Standardization of Whole-Slide Histopathological Images. *IEEE Transactions on Medical Imaging*, 35(2):404–415, 2016.

[25] B. E. Bejnordi, M. Balkenhol, G. Litjens, R. Holland, P. Bult, N. Karssemeijer and J. A. W. M. van der Laak. Automated Detection of DCIS in Whole-Slide H&E Stained Breast Histopathology Images. *IEEE Transactions on Medical Imaging*, 35(9):2141–2150, 2016.

[26] B. Pang, Y. Zhang, Q. Chen, Z. Gao, Q. Peng and X. You. Cell Nucleus Segmentation in Color Histopathological Imagery Using Convolutional Networks. In *Proceedings of Chinese Conference on Pattern Recognition (CCPR)*, pages 1–5. IEEE, 2010.

[27] C. Cong et al. Texture Enhanced Generative Adversarial Network For Stain Normalisation In Histopathology Images. In *IEEE International Symposium on Biomedical Imaging (ISBI)*, pages 1949–1952, 2021.

[28] C. Genestie et al. Comparison of The Prognostic Value of Scarff-Bloom-Richardson and Nottingham Histological Grades in A Series of 825 Cases of Breast Cancer: Major Importance of The Mitotic Count As A Component of Both Grading Systems. *Anticancer Research*, 18(1B):571–576, 1998.

[29] C.-S. Chang et al. Cell Segmentation Algorithm Using Double Thresholding with Morphology-Based Techniques. In *Proceedings of IEEE International Conference on Consumer Electronics, Taiwan*, pages 1–5, 2018.

[30] C. Villani. *Optimal Transport: Old and New.* Grundlehren Der Mathematischen Wissenschaften. Springer Berlin Heidelberg, 2008.

[31] D. A. V. Valen et al. Deep Learning Automates The Quantitative Analysis of Individual Cells in Live-cell Imaging Experiments. *PLoS Computational Biology*, 12(11):e1005177, 2016.

[32] D. Anand et al. Fast GPU-Enabled Color Normalization for Digital Pathology. In *International Conference on Systems, Signals and Image Processing (IWSSIP)*, pages 219–224, 2019.

[33] D. L. Cohn. *Measure Theory.* Birkhäuser Advanced Texts Basler Lehrbücher. Springer New York, 2013.

[34] D. L. Ruderman, T. W. Cronin and C.-C. Chiao. Statistics of Cone Responses to Natural Images: Implications for Visual Coding. *Journal of the Optical Society of America A*, 15(8):2036–2045, 1998.

[35] D. Magee, D. Treanor, D. Crellin, M. Shires, K. Smith, K. Mohee and P. Quirke. Colour Normalisation in Digital Histopathology Images. In *Proceedings of Optical Tissue Image analysis in Microscopy, Histopathology and Endoscopy (MICCAI Workshop)*, volume 100, pages 100–111. Daniel Elson London, 2009.

149

[36] E. D. Gelasca, J. Byun, B. Obara and B. S. Manjunath. Evaluation and Benchmark for Biological Image Segmentation. In *Proceedings of IEEE International Conference on Image Processing*, pages 1816–1819, 2008.

[37] E. Meijering. Cell Segmentation: 50 Years Down the Road [Life Sciences]. *IEEE Signal Processing Magazine*, 29(5):140–145, 2012.

[38] E. Reinhard, M. Adhikhmin, B. Gooch and P. Shirley. Color Transfer Between Images. *IEEE Computer Graphics and Applications*, 21(5):34–41, 2001.

[39] D. Mahapatra et al. Structure Preserving Stain Normalization of Histopathology Images Using Self Supervised Semantic Guidance. In *23rd International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI)*, pages 309–âĂŞ319, 2020.

[40] O. Ronneberger et al. U-Net: Convolutional Networks for Biomedical Image Segmentation. In *Proceedings of Medical Image Computing and Computer-Assisted Intervention*, pages 234–241, 2015.

[41] F. G. Zanjani, S. Zinger, B. E. Bejnordi, J. A. W. M. van der Laak and P. H. N. de With. Stain Normalization of Histopathology Images Using Generative Adversarial Networks. In *Proceedings of IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*, pages 573–577, 2018.

[42] F. Ghaznavi, A. Evans, A. Madabhushi and M. Feldman. Digital Imaging in Pathology: Whole-Slide Imaging and Beyond. *Annual Review of Pathology: Mechanisms of Disease*, 8(1):331–359, 2013.

[43] F. Mahmood et al. Deep Adversarial Training for Multi-Organ Nuclei Segmentation in Histopathology Images. *IEEE Transactions on Medical Imaging*, 39(11):3257–3267, 2020.

[44] G. D. Finlayson, S. D. Hordley and P. M. HubeL. Color by Correlation: A Simple, Unifying Framework for Color Constancy. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(11):1209–1221, 2001.

[45] G. Peyré and M. Cuturi. *Computational Optimal Transport: With Applications to Data Science*. Foundations and Trends in Machine Learning. Now Foundations and Trends, 2019.

[46] H. Chen et al. DCAN: Deep Contour-Aware Networks for Object Instance Segmentation from Histology Images. *Medical Image Analysis*, 36:135–146, 2017.

[47] H. Cho, S. Lim, G. Choi and H. Min. Neural Stain-style Transfer Learning Using GAN for Histopathological Images. *arXiv preprint arXiv:1710.08543*, 2017.

[48] H. Kang, D. Luo, W. Feng, S. Zeng, T. Quan, J. Hu and X. Liu. Stainnet: A Fast and Robust Stain Normalization Network. *Frontiers in Medicine*, 8:746307, 2021.

[49] H. Liang, K. N. Plataniotis and X. Li. Stain Style Transfer of Histopathology Images via Structure-Preserved Generative Learning. In *Proceedings of Machine Learning for Medical Image Reconstruction: Third International Workshop, MLMIR 2020, Held in Conjunction with MICCAI 2020, Lima, Peru, October 8, 2020, Proceedings 3*, pages 153–162. Springer, 2020.

[50] H. Nishar, N. Chavanke and N. Singhal. Histopathological Stain Transfer Using Style Transfer Network with Adversarial Loss. In *Proceedings of Medical Image Computing and Computer Assisted Intervention–MICCAI 2020: 23rd International Conference, Lima, Peru*, pages 330–340. Springer, 2020.

[51] H. Noh, S. Hong and B. Han. Learning Deconvolution Network for Semantic Segmentation. In *Proceedings of the IEEE international conference on computer vision*, pages 1520–1528, 2015.

[52] H. Oda et al. BESNet: Boundary-Enhanced Segmentation of Cells in Histopathological Images. In *Medical Image Computing and Computer Assisted Intervention–MICCAI 2018: 21st International Conference, Granada, Spain, September 16-20, 2018, Proceedings, Part II 11*, pages 228–236. Springer, 2018.

[53] H. Qu et al. Nuclei Segmentation Using Mixed Points and Masks Selected From Uncertainty. In *Proceedings of IEEE International Symposium on Biomedical Imaging*, pages 973–976, 2020.

[54] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville and Y. Bengio. Generative Adversarial Networks. *Communications of the ACM*, 63(11):139–144, 2020.

[55] J. G. Elmore et al. Diagnostic Concordance Among Pathologists Interpreting Breast Biopsy Specimens. *The Journal of the American Medical Association*, 313(11):1122–1132, 2015.

[56] J. I. Epstein, W. C. Allsbrook Jr., M. B. Amin and L. L. Egevad. Update on the Gleason Grading System for Prostate Cancer: Results of An International Consensus Conference of Urologic Pathologists. *Advances in Anatomic Pathology*, 13(1):57–59, 2006.

[57] J. Long, E. Shelhamer and T. Darrell. Fully Convolutional Networks for Semantic Segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3431–3440, 2015.

[58] J. M. Tomczak. General Invertible Transformations for Flow-based Generative Modeling. In *Proceedings of ICML Workshop on Invertible Neural Networks, Normalizing Flows, and Explicit Likelihood Models*, 2021.

[59] J. Newberg and R. F. Murphy. A Framework for the Automated Analysis of Subcellular Patterns in Human Protein Atlas Images. *Journal of Proteome Research*, 7(6):2300–2308, 2008.

[60] J. Song et al. Multi-Layer Boosting Sparse Convolutional Model for Generalized Nuclear Segmentation from Histopathology Images. *Knowledge-Based Systems*, 176:40–53, 2019.

[61] J. Song, L. Xiao and Z. Lian. Contour-Seed Pairs Learning-Based Framework for Simultaneously Detecting and Segmenting Various Overlapping Cells/Nuclei in Microscopy Images. *IEEE Transactions on Image Processing*, 27(12):5759–5774, 2018.

[62] J. T. Kwak et al. Nucleus Detection Using Gradient Orientation Information and Linear Least Squares Regression. In *Proceedings of Medical Imaging: Digital Pathology*, volume 9420, pages 152–159, 2015.

[63] J. Xu, L. Xiang, G. Wang, S. Ganesan, M. Feldman, N. NC Shih, H. Gilmore and A. Madabhushi. Sparse Non-negative Matrix Factorization (SNMF) Based Color Unmixing for Breast Histopathological Image Analysis. *Computerized Medical Imaging and Graphics*, 46:20–29, 2015.

[64] J.-Y. Zhu et al. Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks. In *Proceedings of IEEE International Conference on Computer Vision (ICCV)*, pages 2242–2251, 2017.

[65] K. Chen, N. Zhang, L. Powers and J. Roveda. Cell Nuclei Detection and Segmentation for Computational Pathology Using Deep Learning. In *Proceedings of Spring Simulation Conference (SpringSim)*, pages 1–6. IEEE, 2019.

[66] K. He et al. Mask R-CNN. In *Proceedings of IEEE International Conference on Computer Vision*, pages 2961–2969, 2017.

[67] K. He, X. Zhang, S. Ren and J. Sun. Deep Residual Learning for Image Recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 770–778, 2016.

[68] K. N. Plataniotis and A. N. Venetsanopoulos. *Color Image Processing and Applications.* Springer-Verlag Berlin Heidelberg, New York, USA, 2000.

[69] K. Nguyen, A. Sarkar and A. K. Jain. Structure and Context in Prostatic Gland Segmentation and Classification. In *Proceedings of International Conference on Medical Image Computing and Computer-assisted Intervention*, pages 115–123, 2012.

[70] K. Thierbach, P.-L. Bazin, W. de Back, F. Gavriilidis, E. Kirilina, C. Jäger, M. Morawski, S. Geyer, N. Weiskopf and N. Scherf. Combining Deep Learning and Active Contours Opens The Way to Robust, Automated Analysis of Brain Cytoarchitectonics. In *Proceedings of Machine Learning in Medical Imaging: 9th International Workshop, MLMI 2018, Held in Conjunction with MICCAI 2018, Granada, Spain*, pages 179–187. Springer, 2018.

[71] L. A. Zadeh. Fuzzy Sets. *Information and Control*, 8:338–353, 1965.

[72] L.-C. Chen et al. DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 40(4):834–848, 2018.

[73] L. Dinh, J. S. -Dickstein and S. Bengio. Density Estimation Using Real NVP. In *Proceedings of International Conference on Learning Representations*, 2017.

[74] L. G. Nyul, J. K. Udupa and X. Zhang. New Variants of A Method of MRI Scale Standardization. *IEEE Transactions on Medical Imaging*, 19(2):143–150, 2000.

[75] L. Gorelick, O. Veksler, M. Gaed, J. A. GÃşmez, M. Moussa, G. Bauman, A. Fenster and A. D. Ward. Prostate Histopathology: Learning Tissue Component Histograms for Cancer Detection and Classification. *IEEE Transactions on Medical Imaging*, 32(10):1804–1818, 2013.

[76] L. Peter, D. Mateus, P. Chatelain, N. Schworm, S. Stangl, G. Multhoff and N. Navab. Leveraging Random Forests for Interactive Exploration of Large Histological Images. In *Proceedings of International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 1–8, 2014.

[77] L. Sha, D. Schonfeld and A. Sethi. Color Normalization of Histology Slides Using Graph Regularized Sparse NMF. In *Proceedings of Medical Imaging 2017: Digital Pathology*, volume 10140, pages 249–259. SPIE, 2017.

[78] L. Szilagyi, Z. Benyo, S. M. Szilagyi and H. S. Adam. MR Brain Image Segmentation Using An Enhanced Fuzzy C-Means Algorithm. In *Proceedings of 25th Annual International Conference of IEEE Engineering in Medicine and Biology Society*, pages 724–726, 2003.

[79] W. Liu, Q. He, and X. He. Weakly Supervised Nuclei Segmentation Via Instance Learning. In *Proceedings of IEEE International Symposium on Biomedical Imaging*, pages 1–5, 2022.

[80] M. Everingham et al. The Pascal Visual Object Classes (VOC) Challenge. *International Journal of Computer Vision*, 88:303–338, 2010.

[81] M. Gavrilovic, J. C. Azar, J. Lindblad, C. WÃďhlby, E. Bengtsson, C. Busch and I. B. Carlbom. Blind Color Decomposition of Histological Images. *IEEE Transactions on Medical Imaging*, 32(6):983–994, 2013.

[82] M. Grimaud. New Measure of Contrast: The Dynamics. In *Image Algebra and Morphological Image Processing III*, volume 1769, pages 292–305. SPIE, 1992.

[83] M. Kuse et al. Local Isotropic Phase Symmetry Measure for Detection of Beta Cells and Lymphocytes. *Journal of Pathology Informatics*, 2(2):2, 2012.

[84] M. Macenko, M. Niethammer, J. S. Marron, D. Borland, J. T. Woosley, X. Guan, C. Schmitt and N. E. Thomas. A Method for Normalizing Histology Slides for Quantitative Analysis. In *Proceedings of IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, pages 1107–1110, 2009.

[85] M. Majurski, P. Manescu, S. Padi, N. Schaub, N. Hotaling, C. Simon and P. Bajcsy. Cell Image Segmentation Using Generative Adversarial Networks, Transfer Learning, and Augmentations. In *Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 1114–1122, 2019.

[86] M. Niethammer, D. Borland, J. S. Marron, J. Woosley and N. E. Thomas. Appearance Normalization of Histology Slides. In *Proceedings of International Workshop on Machine Learning in Medical Imaging*, pages 58–66, 2010.

[87] M. R. Prusty et al. Nuclei Segmentation in Histopathology Images Using Structure-Preserving Color Normalization Based Ensemble Deep Learning Frameworks. *Computers, Materials & Continua*, 77(3), 2023.

[88] M. Salvi, N. Michielli and F. Molinari. Stain Color Adaptive Normalization (SCAN) Algorithm: Separation and Standardization of Histological Stains in Digital Pathology. *Computer Methods and Programs in Biomedicine*, 193:105506, 2020.

[89] M. T. McCann, J. Majumdar, C. Peng, C. A. Castro and J. Kovacevic. Algorithm and Benchmark Dataset for Stain Separation in Histology Images. In *Proceedings of IEEE International Conference on Image Processing*, pages 3953–3957, 2014.

[90] M. T. Shaban, C. Baur, N. Navab and S. Albarqouni. StainGAN: Stain Style Transfer for Digital Histological Images. In *Proceedings of IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*, pages 953–956, 2019.

[91] M. Thoma. A Survey of Semantic Segmentation. *arXiv preprint arXiv:1602.06541*, 2016.

[92] M. Veta et al. Automatic Nuclei Segmentation in H&E Stained Breast Cancer Histopathology Images. *Plos One*, 8(7):e70221, 2013.

[93] M. Weigert and U. Schmidt. Nuclei Instance Segmentation and Classification in Histopathology Images with Stardist. In *Proceedings of IEEE International Symposium on Biomedical Imaging Challenges*, pages 1–4, 2022.

[94] Y. Ma, J. Shi, Z. Jiang, and H. Feng. PLSA-Based Pathological Image Retrieval for Breast Cancer with Color Deconvolution. In *Proceedings of MIPPR 2013: Parallel Processing of Images and Optimization and Medical Imaging Processing*, volume 8920, pages 135–141. SPIE, 2013.

[95] N. Hatipoglu and G. Bilgin. Cell Segmentation in Histopathological Images with Deep Learning Algorithms by Utilizing Spatial Relationships. *Medical & biological engineering & computing*, 55:1829–1848, 2017.

[96] N. I. Fisher. *Statistical Analysis of Circular Data*. Cambridge University Press, 1995.

[97] N. Kumar et al. A Dataset and a Technique for Generalized Nuclear Segmentation for Computational Pathology. *IEEE Transactions on Medical Imaging*, 36(7):1550–1560, 2017.

[98] N. L. Johnson et al. *Continuous Univariate Distributions*. vol. 2, John Wiley & Sons, 1995.

[99] N. Otsu. A Threshold Selection Method from Gray-Level Histograms. *IEEE Transactions on Systems, Man, and Cybernetics*, 9(1):62–66, 1979.

[100] O. Martos et al. Optimized Detection and Segmentation of Nuclei in Gastric Cancer Images Using Stain Normalization and Blurred Artifact Removal. *Pathology - Research and Practice*, 248:154694, 2023.

[101] O. Russakovsky et al. ImageNet Large Scale Visual Recognition Challenge. *International Journal of Computer Vision*, 115:211–252, 2015.

[102] P. Filipczuk et al. Computer-Aided Breast Cancer Diagnosis Based on the Analysis of Cytological Images of Fine Needle Biopsies. *IEEE Transactions on Medical Imaging*, 32(12):2169–2178, 2013.

[103] P. J. Rousseeuw. Silhouettes: A Graphical Aid to the Interpretation and Validation of Cluster Analysis. *Journal of Computational and Applied Mathematics*, 20:53–65, 1987.

[104] P. Maji and S. Mahapatra. Rough-Fuzzy Circular Clustering for Color Normalization of Histological Images. *Fundamenta Informaticae*, 164(1):103–117, 2019.

[105] P. Maji and S. Mahapatra. Circular Clustering in Fuzzy Approximation Spaces for Color Normalization of Histological Images. *IEEE Transactions on Medical Imaging*, 39(5):1735–1745, 2020.

[106] P. Naylor, M. Lae, F. Reyal and T. Walter. Nuclei Segmentation in Histopathology Images Using Deep Neural Networks. In *Proceedings of IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017)*, pages 933–936, 2017.

[107] P. Naylor, M. Lae, F. Reyal and T. Walter. Segmentation of Nuclei in Histopathology Images by Deep Regression of the Distance Map. *IEEE Transactions on Medical Imaging*, 38(2):448–459, 2019.

[108] D. M. Pelt and J. A Sethian. A Mixed-Scale Dense Convolutional Neural Network for Image Analysis. *Proceedings of the National Academy of Sciences*, 115(2):254–259, 2018.

[109] Q. Jin et al. RA-UNet: A Hybrid Deep Attention-Aware Network to Extract Liver and Tumor in CT Scans. *Frontiers in Bioengineering and Biotechnology*, 8:605132, 2020.

[110] Q. Wang, S. Wang, X. Zhu, T. Liu, Z. Humphrey, V. Ghukasyan, M. Conway, E. Scott, G. Fragola, K. Bradford and others. Accurate and High Throughput Cell Segmentation Method for Mouse Brain Nuclei Using Cascaded Convolutional Neural Network. In *Proceedings of Patch-Based Techniques in Medical Imaging: Third International Workshop, Patch-MI 2017, Held in Conjunction with MICCAI 2017, Quebec City, QC, Canada*, pages 55–62. Springer, 2017.

[111] R. Achanta et al. SLIC Superpixels Compared to State-of-the-Art Superpixel Methods. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(11):2274–2282, 2012.

[112] S. Ali and A. Madabhushi. An Integrated Region-, Boundary-, Shape-Based Active Contour for Multiple Object Overlap Resolution in Histological Imagery. *IEEE Transactions on Medical Imaging*, 31(7):1448–1460, 2012.

[113] S. E. A. Raza et al. Micro-Net: A Unified Model for Segmentation of Various Objects in Microscopy Images. *Medical Image Analysis*, 52:160–173, 2019.

[114] S. Graham et al. Hover-Net: Simultaneous Segmentation and Classification of Nuclei in Multi-Tissue Histology Images. *Medical Image Analysis*, 58:101563, 2019.

[115] S. Graham et al. Conic: Colon Nuclei Identification and Counting challenge 2022. *arXiv preprint arXiv:2111.14485*, 2021.

[116] S. Kothari, J. H. Phan, R. A. Moffitt, T. H. Stokes, S. E. Hassberger, Q. Chaudry, A. N. Young and M. D. Wang. Automatic Batch-invariant Color Segmentation of Histological Cancer Images. In *Proceedings of IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, pages 657–660, 2011.

[117] S. Liu et al. Unpaired Stain Transfer Using Pathology-Consistent Constrained Generative Adversarial Networks. *IEEE Transactions on Medical Imaging*, 40(8):1977–1989, 2021.

[118] S. Mahapatra and P. Maji. Truncated Normal Mixture Prior Based Deep Latent Model for Color Normalization of Histology Images. *IEEE Transactions on Medical Imaging*, 42(6):1746–1757, 2023.

[119] S. Roy, S. Lal and J. R. Kini. Novel Color Normalization Method for Hematoxylin & Eosin Stained Histopathology Images. *IEEE Access*, 7:28982–28998, 2019.

[120] S. Sornapudi et al. Deep Learning Nuclei Detection in Digitized Histology Images by Superpixels. *Journal of pathology informatics*, 9(1):5, 2018.

[121] S. Vijh, M. Saraswat and S. Kumar. A new complete color normalization method for h&e stained histopatholgical images. *Applied Intelligence*, 51(11):7735–7748, 2021.

[122] M. Saha and C. Chakraborty. Her2Net: A Deep Framework for Semantic Segmentation and Classification of Cell Membranes and Nuclei in Breast Cancer Evaluation. *IEEE Transactions on Image Processing*, 27(5):2189–2200, 2018.

[123] D. F Swinehart. The Beer-Lambert Law. *Journal of Chemical Education*, 39(7):333, 1962.

[124] T. Kwon and J. C. Ye. Cycle-Free CycleGAN Using Invertible Generator for Unsupervised Low-Dose CT Denoising. *IEEE Transactions on Computational Imaging*, 7:1354–1368, 2021.

[125] T. Wollmann, C. S. Eijkman and K. Rohr. Adversarial Domain Adaptation to Improve Automatic Breast Cancer Grading in Lymph Nodes. In *Proceedings of IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*, pages 582–585, 2018.

[126] T.-Y. Lin et al. Microsoft COCO: Common Objects in Context. In *Proceedings of Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland*, pages 740–755. Springer, 2014.

[127] V. Badrinarayanan, A. Kendall and R. Cipolla. SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(12):2481–2495, 2017.

[128] W. W. Parson. *Modern Optical Spectroscopy*. Springer, 2007.

[129] W. Zhou et al. Image Quality Assessment: From Error Visibility to Structural Similarity. *IEEE Transactions on Image Processing*, 13(4):600–612, 2004.

[130] X. Chen, Y. Duan, R. Houthooft, J. Schulman, I. Sutskever and P. Abbeel. Info-GAN: Interpretable Representation Learning by Information Maximizing Generative Adversarial Nets. In *Proceedings of the 30th International Conference on Neural Information Processing Systems*, pages 2180–2188, 2016.

[131] X. Huang et al. Multimodal Unsupervised Image-to-image Translation. In *Proceedings of European Conference on Computer Vision (ECCV)*, pages 179–196, 2018.

[132] X. Li and K. N. Plataniotis. A Complete Color Normalization Approach to Histopathology Images Using Color Cues Computed from Saturation-Weighted Statistics. *IEEE Transactions on Biomedical Engineering*, 62(7):1862–1873, 2015.

[133] X. Li and K. N. Plataniotis. Blind Stain Decomposition for Histopathology Images Using Circular Nature of Chroma Components. In *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 877–881, 2015.

[134] X. Li and K. Plataniotis. Circular Mixture Modeling of Color Distribution for Blind Stain Separation in Pathology Images. *IEEE Journal of Biomedical and Health Informatics*, 21(1):150–161, 2017.

[135] X. Liu, Z. Deng and Y. Yang. Recent Progress in Semantic Image Segmentation. *Artificial Intelligence Review*, 52:1089–1106, 2019.

[136] X. Mao et al. On the Effectiveness of Least Squares Generative Adversarial Networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 41(12):2947–2960, 2019.

[137] X. Pan et al. Accurate Segmentation of Nuclei in Pathological Images via Sparse Reconstruction and Deep Convolutional Networks. 229:88–99, 2017.

[138] Y. Cui et al. A Deep Learning Algorithm for One-step Contour Aware Nuclei Segmentation of Histopathology Images. *Medical & biological engineering & computing*, 57:2027–2043, 2019.

[139] Y. K. Lai and P. L. Rosin. Efficient Circular Thresholding. *IEEE Transactions on Image Processing*, 23(3):992–1001, 2014.

[140] Y. Lin et al. BoNuS: Boundary Mining for Nuclei Segmentation with Partial Point Labels. *IEEE Transactions on Medical Imaging*, pages 1–11, 2024.

[141] Y. Liu, P. Zhang, Q. Song, A. Li, P. Zhang and Z. Gui. Automatic Segmentation of Cervical Nuclei Based on Deep Learning and a Conditional Random Field. *IEEE Access*, 6:53709–53721, 2018.

[142] Y. Ma, Z. Jiang, H. Zhang, F. Xie, Y. Zheng and H. Shi. Proposing regions from histopathological whole slide image for retrieval using selective search. In *Proceedings of IEEE International Symposium on Biomedical Imaging (ISBI 2017)*, pages 156–159, 2017.

[143] Y. Ma, Z. Jiang, H. Zhang, F. Xie, Y. Zheng, H. Shi, Y. Zhao and J. Shi. Generating Region Proposals for Histopathological Whole Slide Image Retrieval. *Computer Methods and Programs in Biomedicine*, 159:1–10, 2018.

[144] Y. Y. Wang, S. C. Chang, L. W. Wu, S. T. Tsai and Y. N. Sun. A Color-Based Approach for Automated Segmentation in Tumor Tissue Classification. In *Proceedings of Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 6576–6579, 2007.

[145] Y. Zheng et al. Stain Standardization Capsule for Application-Driven Histopathological Image Normalization. *IEEE Journal of Biomedical and Health Informatics*, 25(2):337–347, 2021.

[146] Y. Zheng, Z. Jiang, H. Zhang, F. Xie, Y. Ma, H. Shi and Y. Zhao. Histopathological Whole Slide Image Analysis Using Context-Based CBIR. *IEEE Transactions on Medical Imaging*, 37(7):1641–1652, 2018.

[147] Z. Pawlak. *Rough Sets: Theoretical Aspects of Resoning About Data*. Kluwer, Dordrecht, The Netherlands, 1991.

[148] Z. Qian et al. Transformer Based Multiple Instance Learning for Weakly Supervised Histopathology Image Segmentation. In *Proceedings of Medical Image Computing and Computer Assisted Intervention*, pages 160–170, 2022.

[149] Z. Wang et al. Image Quality Assessment: From Error Visibility to Structural Similarity. *IEEE Transactions on Image Processing*, 13(4):600–612, 2004.

[150] Z. Zhou et al. UNet++: A Nested U-Net Architecture for Medical Image Segmentation. In *Proceedings of Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*, pages 3–11, 2018.

[151] N. Zhou and Y. Gao. Optimized Color Decomposition of Localized Whole Slide Images and Convolutional Neural Network for Intermediate Prostate Cancer Classification. In *Proceedings of Medical Imaging 2017: Digital Pathology*, volume 10140, pages 225–233. SPIE, 2017.