Morphological Network: Network With Morphological Neurons

Ranjan Mondal

Electronics and Communication Sciences Unit

Indian Statistical Institute

Supervisor: Prof. Bhabatosh Chanda

A thesis submitted to Indian Statistical Institute for partial fulfillment of the requirements of the degree of

Doctor of Philosophy in Computer Science

August 2021

To My Parents

Acknowledgments

Abstract

Image processing with traditional approaches mainly use the tools of linear systems. However, linear approaches are not well suited and may even fail to solve problems involving geometrical aspects of the image. Thus, nonlinear geometric approaches like morphological operations are very popular in those cases. Morphological operations are nonlinear operations based on a set and lattice-theoretic methodology for image analysis that are capable of describing the geometrical structure of image objects quantitatively. It is suitable for various problems in image processing, computer vision, and pattern recognition. While solving problems with morphology, a particular structuring element is defined. Structuring elements have particular shape and size which are applied spatially in the images. Finding such structuring elements for each task are very difficult and hand engineered. In this thesis, we develop networks with trainable morphological structuring elements for solving several problems. Our main idea is to learn appropriate structuring element(s) given an objective. The elementary operations of morphology are dilation and erosion. Similar to convolutional neural networks, a network is built with dilation and erosion operators with trainable structuring elements. For example, we have considered a gray scale rainy dataset. Since the rain streak has a particular shape and is considered as white noise, the network is able to remove rain in grayscale images using learned structuring elements. Dilation and Erosion in particular order constitute opening and closing operations. Opening and closing are popular in removing bright and dark noise from images. We have relied more on the training of structuring elements and built a network with dilation and erosion so that it may perform opening or closing operations based on the necessity. We have empirically proved that opening and closing is happening in the network. Further the network is applied for image restoration tasks and evaluated on colour image de-raining and image dehazing. Dilation and Erosion are composed with max and min operation. To make it more generic like a neural network, we have theoretically analyzed the morphological network and have built a dense morphological network to process 1-dimensional feature vectors. Morphological block has been defined by a dilation-erosion layer followed by a linear combination layer. We have shown that a morphological block represents a sum of hinge functions. With this morphological block our network is able to perform many classification tasks. Further, we have proved that two sequential morphological blocks can approximate any continuous function. We have also analyzed the network with deep

multilayer configuration and shown many properties of the network. Next, We have extended the dense morphological concept and built a 2D network so that it can be applied in general image processing tasks. We build a network with a basic 2D morphological block i.e dilation erosion followed by linear combination of feature map. We have repeated this block and built a network for general image processing tasks such as classification of pixels. We have also evaluated the performance of the network on image processing tasks like segmentation of blood vessels from fundus images, segmentation of lungs from chest x-ray and image dehazing.

Contents

Contents

List of Figures

List of Tables

Part I.

Introduction

Chapter 1.

Introduction

1.1. Motivation

Mathematical morphology [83] is a popular non-liner technique, where the operators directly handle with shapes. The operators are defined in terms of probe(s) called *structuring elements*. This technique is successfully applied on various computer vision problems including image restoration, image feature extraction and automated visual inspection. *Dilation* and *erosion* are two elementary operations in morphology, which may be combined to form other useful operators, like *opening* and *closing*.

On the other hand, from the last decade deep learning is a very popular and major area in computer vision. Although it was developed 20 years ago, it becomes more popular in modern times due to development of computer hardware. Convolutional neural network (CNN) is a class of deep neural networks which is very successful in many image processing and classification task. CNN is found to be very useful in extracting image features and solving different image processing problems such as image denoising [105, 104], semantic segmentation [8, 60]) and object detection [26, 28, 68]. Variations of of CNN, like R-CNN, fast R-CNN, faster R-CNN and mask R-CNN are also used to solve different computer vision problems. The basic building block of CNN is convolution of feature map with kernels. Feature map at the entry layer is nothing but the input image.

It maybe noted that structuring element and convolution kernel are analogous. Both of them operate spatially on the image features. However, convolution is a linear operation and it captures linear features. So, with the convolution operation it is very hard to capture nonlinear information in the image like shape and size of the objects in the image. On the other hand, morphological operators are nonlinear and they are generally very useful where shape, size, distance and connectivity are involved in the images. These problems are solved by choosing structuring element(s) of appropriate shape and size intelligently by the user.

Finding customized / tailored size and shape of the structuring elements and also the order in which erosion and dilation operations are to be applied still remain a huge challenge [84].

It is even harder to design the structuring element especially when it comes to gray-scale or color image processing. Thus there is a need to learn the structuring element automatically from the data itself given the objective. Convolutional neural networks are able to extract complex features hierarchically from images as they are used in multiple sequential layers. In this thesis we have built layers with basic morphological operators like dilation and erosion and used sequentially to extract morphological features hierarchically. The structuring elements are learned by back propagation algorithm. A brief overview of the existing works related to morphological network is given below.

1.2. Related Work

Mathematical morphology has a very rich mathematical foundation with set and lattice theoretic methodology for image analysis. It can quantitatively describe the geometrical structure of image objects. In the late 1960's it was started to analyze[81, 80] binary images from geological and biomedical data. Later it was formalized and extended for binary pattern recognition[66, 76] based on cellular automata and Boolean/threshold logic. Morphology was first applied in gray-level images[80] in late 1970's. It was brought to the mainstream of image/signal processing and other nonlinear filtering approaches in mid 1980's. With the help of lattice theory it was later generalized[31, 82] with in 1980's and 1990's. Currently it is employed in the analysis of graphs, meshes and many other spatial structures, although digital images remain its common application domain. Morphological operators decompose objects or shapes into meaningful parts which helps in understanding them in terms of the elements. Since identification of objects and their features are directly correlated with their shapes and arrangement, morphological methods are quite suited for visual tasks [17]. Schonfeld *et al.* has done theoretical analysis of morphological filters [79] for the optimal restoration of noisy binary images. They proved that the class of alternating sequential filters is a set of parametric, smoothing morphological filters that best preserve the crucial structure of input images. In a separate work, mukhopadhyay *et al.* extended traditional contrast enhancement task using mathematical morphology. There scale-specific features of the image is extracted using multi scale tophat transformation and they are modified for achieving local contrast enhancement [56]. In the book of Alan Bovik [11], it is shown that morphology can be used for template matching. In the paper of feehs *et al.*, we introduce a general robust N-dimensional morphological edge detector [22]. It is highly suitable for many shape-oriented problems [6], including character recognition [50]. Mendel *et al.* has shown that, extracting specific morphological attributes from the scanned image, the dynamic OCR system is able to generalize and approximate similar images. Angulo *et al.* applied mathematical morphology operators to quantify the shape of round-objects which present irregularities from an ideal circular patterns in Microscopic cell

images.

Now a days, there are many industrial application of mathematical morphology such as teeth segmentation [77], texture analysis [97], curvature evaluation [101]. Said *et al.* proposed mathematical morphological approach for teeth segmentation from dental radiographic films. They also proposed a grayscale contrast stretching transformation to improve the performance of teeth segmentation. Zana *et al.* presents an algorithm based on mathematical morphology and curvature evaluation for the detection of vessel-like patterns in a noisy environment. xia *et al.* proposed local morphological multifractal exponents to characterize the local scaling properties of textures.

The basic concept of learning of structuring elements was first introduced in 1993 by defining Morphological perceptron. Morphological perceptron was introduced by [17] and the authors used morphological network to solve template identification problem. Later it was generalized by [72] to tackle the problem of binary classification by restricting the network to single layer architecture. The decision boundaries were considered as parallel to the axes. This single-layer architecture has been extended to two-layer architecture by Sussner [89]. This two-layer architecture can learn multiple axis-parallel hyperplanes, and therefore is able to solve arbitrary binary classification task. But, in general the decision boundaries may not be axis-parallel, so, large number of hyperplanes need to be learned by the network if the decision boundaries are not parallel to the axes. To train the structuring elements of the network, researchers [62] have tried to use gradient descent by combining classical perceptron with morphological perceptron. In [1] they used linear function to apply it to regression problems. They have utilized network architecture similar to morphological perceptrons with competitive learning to forecast stock markets. With dendritic structure of morphological neurons, Zamora *et al.* [100] replaced the `argmax` operator by `softmax` function to overcome the problem of gradient computation and used gradient descent to train the network. In [47] authors have given a framework for morphological structuring elements using counter–harmonic mean. Similar to morphological operation [106] defined max-plus operators and Applied for filter selection and model pruning in neural networks.

Recently morphological perceptrons has been renewed in [15] They have training algorithms of morphological perceptrons and analyzed with geometry. Morph-CNN [49] used both convolutional neural network and morphological operators for multi digits recognition. Recently few works [2, 23, 59] has been done in this filed. Nogueira et al.[59] introduces morphological network with multiple morphological operation for image classification task. Islam et.al [2] used morphological hit-or-miss transform to build the network. Limonova et.al [44] proposed bipolar morphological neural networks to do convolution without multiplication. In very recent work Shih et.al [85] proposed a deep learning framework to learn desired structuring element from data.

1.3. Computer vision problems addressed with proposed networks

In this manuscript we have build a network using basic morphological operations. With in help of morphological network we have tried to solve many image processing task which are image de-raining image de-hazing, retinal image segmentation, lungs segmentation, old document image binarization, aerial image classification. Each problem have different characteristics which we can measure the efficacy of the network. Image de-raining is a problem where we have to remove rain from a given color images. Rain drops have particular shape and size. There may have different rain patterns in different image instances. Morphological operations are very sensitive to such structural shape of rain drops. On the other hand Image de-hazing is popular image processing task in recent days. We may consider the haze as a white noise and try to remove them using morphological network. Although here we have used morphological network to estimate the amount of haze in the images. Old document image binarization a another problem where the given image has a lot of colored noise and the objective is to classify each pixels into text or non text. Further we have tried to do segmentation using morphological network. We have done segmentation on retinal image and chest X-Ray images. Retinal image segmentation the task is to segment blood vessels from the retinal images. It maybe considered as a hard task since the blood vessels are very thin image retinal images. Unlike blood vessels segmentation from the retinal images, lungs segmentation from chest ray images is completely different task. The segmented blob size is large in case of lungs segmentation from chest X-Xray images. With the help of 2D morphological network we have tried to do image classification from aerial images. In aerial image contains a lot of structures, hence morphological operations are very efficient in these cases.

In the next section we have highlighted our contribution in this thesis.

1.4. Contribution

In this thesis, we develop networks with trainable morphological structuring elements for solving various problems. Our main idea is to learn appropriate structuring element(s) given an objective. Before we get into it we have to verify whether it is able to learn the structuring element or not.

1.4.1. Automatic learning of structuring elements

We proposes that the structuring element can be learned using back-propagation algorithm. Similar to CNN, a network is built with dilation and erosion operators with trainable
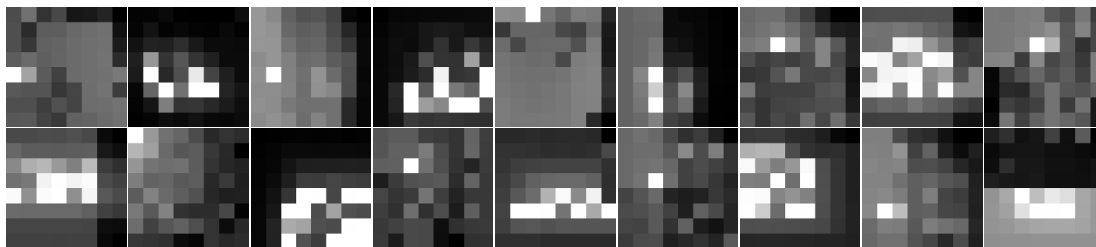
Figure 1.1.: Learned structuring elements at different layers of a morphological network. The most bright pixel is displayed by the maximum value of the structuring elements and the darkest pixel by minimum value.

structuring elements. We have verified the claim by experimenting with gray scale rainy dataset. Since the rain streak has particular shape and is considered as white noise, the network is able to remove rain in grayscale images using learned structuring elements. We observe that a single morphological network can de-rain an image with any arbitrary shaped rain-droplets and achieves similar performance with the contemporary CNNs. It is interesting to note that the proposed morphological network needs much less number of trainable parameters (network size) compared to the said CNN. Figure 1.1 shows some example structuring elements used in the network learned by back-propagation algorithm for image de-raining problem.

1.4.2. Morphological Opening-Closing Network

Dilation and Erosion in particular order constitute opening and closing operations. We have relied more on the training of structuring elements and built a network with dilation and erosion so that it may perform opening or closing operation based on the necessity. Opening operation means erosion followed by dilation with a same structuring element. In this chapter, we have shown that dilating and eroding with different structuring element may also produce same output as opening under some constraints. Similarly it follows for closing operation. We have empirically proved that opening and closing is happening in the network. The network is applied for image restoration task and evaluated on colour image de-raining and image de-hazing.

1.4.3. Dense Morphological Network

Dilation and Erosion are composed with max and min operation. We have theoretically analyzed the dense morphological network in detail. We build a dense morphological network to process 1-dimensional feature vectors. We have shown that a dilation-erosion layer followed by a linear combination layer represents a sum of hinge functions. With this configuration it is able to perform many classification task. Here we have also analyzed

the network with deep multilayer configuration. We have also shown many properties of the network. To experimentally validate the efficacy of dense morphological network in real-life applications, we have evaluated its performance on standard classification datasets like MNIST, Fashion-MNIST, SVHN, CIFAR-10.

1.4.4. 2D morphological Network

Next, We have extended the dense morphological concept and build a 2D network for many image processing tasks. Here we have extended the idea of dense morphological network for general image processing task. We build a network with a basic morphological block i.e dilation erosion followed by linear combination of feature map. We have repeated this block and build a network for general image processing task. Specially classification of pixel. We have evaluated the performance of the network on image processing task like segmentation of blood vessel from fundus images, segmentation of lungs from chest x-ray and image dehazing.

1.5. Organization of Thesis

This thesis contains six chapters out of which Chapters 2 to 5 are contributory chapters. First we have presented process of automatic learning of structuring elements [55] in chapter **??**. There we have shown that sequence structuring of element can be learned using back-propagation algorithm. We have shown this with an example of gray-scale image de-raining task.

In chapter 2, we have built a network with dilation and erosion operators on successive layers so that it can preform opening and closing operations [52]. Here we have shown that it is possible to perform like opening or closing with dilation and erosion operations even using different structuring elements. On the basis of this observation we build a network so that opening or closing may happen in the network using trained structuring, which may not be same. The network is applied for colour image draining and image de-hazing.

Next in chapter 3 we have analyzed the dense morphological network [54] in details. We have shown that a dilation-erosion layer followed by a linear combination layer represents a sum of hindge functions. With this configuration we are able to perform many classification tasks. This leads to network with deep multilayer configuration.

Next we have extended the idea of dense morphological network for general image processing tasks[52, 54] including classification tasks. This is discussed in chapter 4 with example of segmentation of blood vessel from fundus images, segmentation of lungs from chest x-ray and image dehazing and old document image binarization.

Finally, we have concluded this thesis in chapter 5 by discussing the issues addressed in

the previous chapters and also outline the possible future directions where the effort needs to be focused one further progress the state-of-the art in terms of both - methodologies and applications.

Part II.

Image Restoration by Opening Closing Network

Chapter 2.

Image Restoration by Opening-Closing Network

2.1. Introduction

Morphological operations are very effective in solving various challenging problems [75, 66] involving shapes. For the task at hand, the choice and size of the Structuring Element (SE) greatly affect the performance of the methods, apart from the choice of the morphological operators [19]. Also, a single operator seldom suffices to obtain the desired results. Usually, different operators need to be applied in a particular sequence. Choosing correct sequence of correct operators along with associated SEs can be extremely hard in practice.

In the previous chapter we have shown that a sequence of dilation and erosion can be learned using back propagation algorithm for gray-scale image de-raining. In this chapter, we will try to analyze the sequence of dilation and erosion and the working of the network. We propose and implement morphological network replicating the effect of opening and closing (a combination of dilation and erosion) through automatic learning of SEs. We have shown that the network can perform opening and closing operation using the trained structuring element. We have also modified the network into an opening-closing network for general image restoration task. To demonstrate the effectiveness of our approach, we have applied our proposed network on two image restoration problems: image de-raining and image de-hazing.

For image de-raining, Luo *et al.* [45] assumed sparse rain streaks having similar orientations. They proposed a discriminative approach that approximated the clean background and removed rain streak components. Chen *et al.* used a low patch-rank prior to capture rain pattern. Patch-based GMM priors was proposed by Li *et al.* [43] to model rain streak and background by decomposing input image. Fu *et al.* [24] used CNN to learn the mapping between rainy image and clean image directly. Fu *et al.* [25] decomposed input image into background and rain in different layer, from which CNN is used to remove rain streaks. Considering rain-drops as a white noise, with paricular shape and size we have

used opening-closong network to remove the rain from images.

Image de-hazing is an ill-posed problems, considering that one needs to accurately determine the depth of (partially) obscured surface as well as restore color. He *et al.* [30] proposed dark channel prior method, where the statistic of the outdoor haze-free image was used to find the minimum value in any one of the colour channels. This information was used to determine the haze depth and transmittance and produced the de-hazed image. Ancuti *et al.* [4] used a fast method to estimate transmittance and airlight by identifying hazy region based on the difference in the hue of image and its inverse. Fattal [21] used the concept of color line prior. *Tang et al.* [90] and *Cai et al.* [13] learned the mapping between the hand-crafted features and transmittance. Ren *et al.* [70] proposed CNN network at multiple scales to extract features and to establish their mapping to transmittance. Li *et al.* [42] used a modified haze equation where the parameters were unified to a single variable, which was estimated by a CNN. In our method we have modified the haze equation and tried to map transmittance and airlight from the images using opening-closing network. We do not use any highly optimized or highly complex architecture for the tasks to demonstrate versatility of the proposed network. Rest of the paper is organized as follows.

The proposed method is presented and explained in Section 2.2, where we have defined opening and closing layers with the help of dilation and erosion. We have shown that the learned structuring element can perform opening and closing operation inside the network. After that, we have presented experimental results with discussion on de-raining and de-hazing problems in Section 2.3.

2.2. Proposed Method

Convolution operation and elementary morphological operations such as dilation and erosion are all neighbourhood operations defined in terms of the kernel. However, morphological operators are non-linear operators, while convolution operator is linear. Though both types of operators can be defined on n-dimensions, in this work we consider morphological operations only on two-dimensional (2D) images. In this section, we build layers using dilation and erosion operators. The networks built using the dilation and erosion layers are subsequently applied for de-raining and de-hazing the images.

In mathematical morphology, the objects or its parts in an image are considered to be sets, and the operators are defined in terms of set theoretic translation, union, intersection and complementation to examine/extract various properties of these objects, such as shape and texture, using small geometrical probe known as structuring element (SE). In binary images, the pixels belong to either foreground or background (i.e., complement of foreground) set. These sets are viewed as sets in 2D space. Similarly, a grayscale image X defined over domain $\mathcal{D} = \{(x, y)\}$ may be represented as a 3D set $U_X = \{(x, y, v) \mid v \leq X(x, y)\}$, called

umbra of X. Structuring element W defined over $S = \{(x, y)\}$ may also be represented by a 3D set U_W. Dilation of U_X by U_W is defined as union of all U_X translated to every pixels of U_W. To get back dilated grayscale image top surface of the dilation result is extracted. We may get eroded grayscale image in similar manner. However, a concise and more straightforward formulation of grayscale dilation and erosion is proposed by Sternberg [88] as follows.

2.2.1. Morphological dilation and erosion layers

Let gray-scale image X is of size $M \times N$. The Dilation ($\oplus$) and Erosion ($\ominus$) operations on X are defined, respectively, as [88]

$$(X \oplus W_d)(x, y) = \max_{(l,m) \in S} (X(x - l, y - m) + W_d(l, m)) \tag{2.1}$$

$$(X \ominus W_e)(x, y) = \min_{(l,m) \in S} (X(x + l, y + m) - W_e(l, m)) \tag{2.2}$$

where $W_d(x, y) \in \mathbb{R}$ and $W_e(x, y) \in \mathbb{R}$ are the structuring elements of dilation and erosion operators respectively defined on domain $S = \{ (l, m) \mid l \in \{1, 2, 3, .., a\}; \; m \in \{1, 2, 3, .., b\}\}$. Note that actual shape of the geometric probe (SE) may not always cover entire rectangular domain S, then $W(x, y) = \begin{cases} \geq 0 & (x, y) \in G \\ -\infty & \text{otherwise} \end{cases}$ During implementation a large number is used in place of ∞, and that number is at least 1 (one) plus the maximum possible pixel value in gray scale image. Both dilation and erosion are many-to-one mappings. That means

$$(X \oplus W_1)(x, y) = (X \oplus W_2)(x, y) \quad \text{for some } W_1 \text{ and } W_2 \tag{2.3}$$

may not imply $W_1(x, y) = W_2(x, y)$. For example, let C denotes a curve whose length is greater than the diameter of a disk SE D. Then it can be shown that

$$C \oplus D = C \oplus \delta D$$

where δD denotes the boundary of D. In discrete domain, connectivity of C and D should be chosen appropriately (e.g., if C is 8-connected, then D should be 4-connected and vice versa). Similarly,

$$(X \ominus W_1)(x, y) = (X \ominus W_2)(x, y) \quad \text{for some } W_1 \text{ and } W_2 \tag{2.4}$$

may not imply $W_1(x, y) = W_2(x, y)$. For example, let A be a simple blob (i.e., a connected component without any hole) and D is a disk, then it can be shown that

$$A \ominus D = A \ominus \delta D$$

The examples (as depicted in equations equation 2.3 and equation 2.4) can also be verified by taking the SE as $\delta D \cup D'$, where $D' \subset D$. This notion is useful in learning SEs for compound operators, like opening and closing.

Now, similar to convolution layer, morphological layers can be formed using dilation and erosion operators. We call a layer with dilation (resp. erosion) operation as *dilation-layer* (resp. *erosion-layer*). We also define the output of dilation-layer as *dilation feature map* and that of erosion-layer as *erosion feature map*. Note that at each layer multiple dilation or erosion may be applied using different SEs. Applying k dilation operation will generate k dilation feature map in the next layer. We denote dilation applied on the input with k number of SE of size $a \times b$ by $D_{a \times b}^k$. Similarly, erosion operation by k number of structuring element is denoted by $E_{a \times b}^k$. Note that both dilation and erosion are increasing operations and also are dual to each other.

2.2.2. Morphological opening and closing layers

Elementary morphological operations, i.e., dilation and erosion, are applied in many image processing tasks, such as edge detection. However, opening and closing operations are far more useful and important as these are filters. Being filters these operations satisfy essential properties like increasing and idempotent. Moreover, opening is anti-extensive, while closing is extensive. Like dilation and erosion, opening and closing are also dual to each other. In fact, many operations that are increasing, idempotent and anti-extensive are termed as 'opening'. Some examples are area opening [91] and path opening [32] which follows the properties. Traditionally, opening and closing are defined as a compound operator by concatenating dilation and erosion operators. Thus, these operations are defined in terms of SEs as

$$\text{Opening:} \quad (X \circ W_o)(x, y) = ((X \ominus W_o) \oplus W_o)(x, y) \tag{2.5}$$

$$\text{Closing:} \quad (X \bullet W_c)(x, y) = ((X \oplus W_c) \ominus W_c)(x, y) \tag{2.6}$$

So, morphological opening or closing network can be constructed by cascading morphological dilation and erosion layers as defined in equations equation 2.5 and equation 2.6. Equations equation 2.3 and equation 2.4 suggest that $X \circ W_o$ may be resulted in by dilation

and erosion with SEs different from W_o. In other words, for a class of problems

$$(X \circ W_o)(x, y) = ((X \ominus W_o') \oplus W_o'')(x, y) \tag{2.7}$$

where W_o, W_o' and W_o'' may not be equal to each other. Thus, we may call $((X \ominus W_1) \oplus W_2)(x, y)$ (for some W_1 and W_2) an opening operation if it is

1. Increasing: $X \subset Y$ implies $((X \ominus W_1) \oplus W_2)(x, y) \subseteq ((Y \ominus W_1) \oplus W_2)(x, y)$
2. Idempotent: $((((X \ominus W_1) \oplus W_2) \ominus W_1) \oplus W_2)(x, y) = ((X \ominus W_1) \oplus W_2)(x, y)$
3. Anti-extensive: $((X \ominus W_1) \oplus W_2)(x, y) \subseteq X$

Similar argument is true for closing operation also. Note that to develop opening (or closing) network we could have used and trained the same SEs for dilation and erosion layers following equation equation 2.5 (resp. equation equation 2.6). In that case we need to specify the use of opening or closing networks and there order beforehand to solve the given problem. In this work we have trained the SEs in a sequence of dilation and erosion layers independently so that the opening/closing networks (consequently, alternate sequential filters) evolve along with appropriate SEs to solve the problem in hand. Moreover, multiple dilation/erosion operations with different SEs W_k ($k = 1, 2, 3, \cdots$) may be applied on an image or on already computed feature map to produce multiple dilation or erosion feature maps. These multiple feature maps are expected to highlight different types of features in the image based on the profile of the SEs.

All the SEs are initialized randomly while building the network and are trained using back-propagation based on the training samples (images). It may be noted that the erosion and dilation operations use min and max operations, respectively. So, though not fully differentiable, these are at least piece-wise differentiable. Hence, there occurs no problem in back-propagating the gradient to update the SEs. Next we briefly present how back-propagation takes place in the morphological network to train the SEs.

2.2.3. Back-propagation in Morphological Network

Back-propagation algorithm is used to update the SEs in the morphological opening-closing network. We have already stated that our opening and closing networks are built using dilation and erosion layers. Second, dilation and erosion are dual operations, so describing the training of SEs for either of them is sufficient to describe the other. For simplicity, we consider here a single dilation layer. The propagation of gradient thorough the network is very similar to that of a neural network. To start with we recall the expression of the gradient. Suppose an input image X having a dimension of $M \times N \times c$ is passed through a dilation layer that produces an output feature map Y. The structuring element S in the dilation layer has size $A \times B \times c$ and, with appropriate padding, the size of Y is $M \times N$.

The output at location (x, y) of Y can be expressed as :

$$Y(x, y) = \max_{l,m,n}\{X(x + l, y + m, n) + S(l, m, n)\} \tag{2.8}$$

where (l, m, n) denotes the index of the SE S. The SE is trained using back-propagation so that Y approaches desired dilation feature map $\tilde{Y}$. Let L be the loss or error between output of the dilation layer and the desired dilation feature map, i.e.,

$$L = L(\tilde{Y}, Y) \tag{2.9}$$

When SE S is properly trained, $Y \to \tilde{Y}$ which implies $L \to 0$. It is evident from equations equation 2.8 and equation 2.9 that Y and, consequently, L depend on SE S. Using chain rule due to partial derivatives we have

$$\begin{aligned}
\frac{\partial L}{\partial S(x, y, z)} &= \sum_x \sum_y \frac{\partial Y(x, y)}{\partial S(x, y, z)} \frac{\partial L}{\partial Y(x, y)} \\
&= \sum_x \sum_y \nabla Y(x, y) \frac{\partial L}{\partial Y(x, y)}
\end{aligned}$$

where

$$\nabla Y(x, y) = \begin{cases} 1 & \text{if } Y(x, y) = X(x + \delta x, y + \delta y, \delta z) + S(\delta x, \delta y, \delta z) \\ 0, & \text{otherwise} \end{cases} \tag{2.10}$$

Thus the structuring element is updated as

$$S(x, y, z) = S(x, y, z) - \alpha \frac{\partial L}{\partial S(x, y, z)} \tag{2.11}$$

where α is the learning rate. If multiple SEs $S_k(x, y, z)$ $(k = 1, 2, 3, \cdots)$ are used, multiple dilation feature maps Y_k $(k = 1, 2, 3, \cdots)$ are obtained, but the back-propagation strategy would be same.

The idea is straightaway extended to multi-layer network. Suppose an intermediate feature map denoted by X_i is passed through i-th dilation layer and produces an output feature map $X_{i+1,k}$ after dilating with SE $S_{i,k}$ in i-th layer. Then, keeping all other parameters same, equation equation 2.8 can be re-written as

$$X_{i+1,k}(x, y) = \max_{l,m,n}\{X_i(x + l, y + m, n) + S_{i,k}(l, m, n)\} \tag{2.12}$$

Let L is the final loss of the network by comparing desired output and the predicted output.

As before we can calculate the gradient $\dfrac{\partial L}{\partial S_k(x,y,z)}$ with respect to the k-th SE by the following equation using chain rule:

$$
\begin{aligned}
\frac{\partial L}{\partial S_{i,k}(x,y,z)} &= \sum_x \sum_y \frac{\partial X_{i+1,k}(x,y)}{\partial S_{i,k}(x,y,z)} \frac{\partial L}{\partial X_{i+1,k}(x,y)} \\
&= \sum_x \sum_y \nabla X_{i+1,k}(x,y) \frac{\partial L}{\partial X_{i+1,k}(x,y)}
\end{aligned}
$$

where

$$
\nabla X_{i+1,k}(x,y) = \begin{cases} 1, & \text{if } X_{i+1,k}(x,y) = X_i(x+\delta x, y+\delta y, \delta z) + S_k(\delta x, \delta y, \delta z) \\ 0, & \text{otherwise} \end{cases} \tag{2.13}
$$

The term $\dfrac{\partial L}{\partial X_{i+1,k}(x,y)}$ can be obtained by computing recursively starting from the final layer. Finally, k-th SE $S_{i,k}(x,y,z)$ is updated using equation equation 2.11. Similarly, the gradient for the erosion layer can be derived. A worked-out example of gradient calculation for the erosion layer is shown in [23].

2.2.4. Verifying the opening and closing operation

In order to establish the efficacy of our concept of learning morphological opening and closing networks along with associated structuring elements (kernels), we have carried out some initial experiments before applying to real-life problems. However, we first need to show experimentally that the opening network realized by concatenating erosion and dilation layers with not necessarily the same SEs may satisfy desired properties of opening operation, i.e., increasing, idempotent and anti-extensive. In other words, a network made of an erosion layer followed a dilation layer having different SEs satisfies essential properties of opening and, thus, can work like opening network.

Verifying properties of opening operation

First, for training and testing the network to be developed, we have taken 400 natural images from flickr [63] and converted those to grayscale images of size 416×416. Each image is opened with a disk SE of radius 7 to generate the groundtruth of opened image. Second, we construct an opening network by concatenating an erosion layer followed by a dilation layer (see Figure 2.1(a)), where each layer has its own structuring element trained by back-propagation to minimize the loss. We take SEs of size 20×20 and initialize with random numbers from uniform distribution. Suppose image X is input to the network and at the output we obtain image $\gamma(X)$.

(a) A simple Opening Network

(b) Learned Structuring Elements
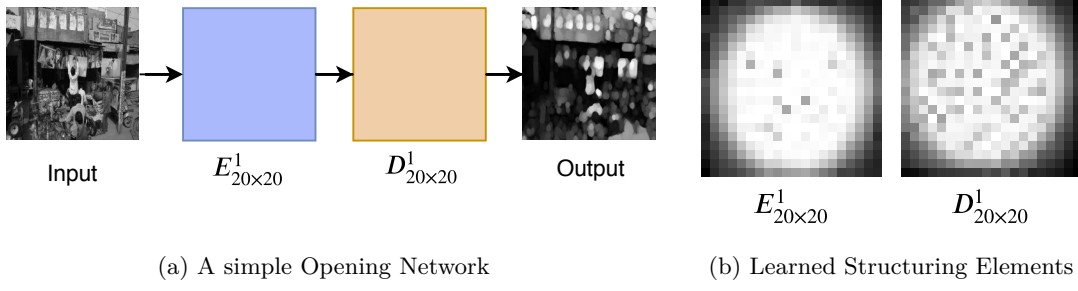
Figure 2.1.: (a) A simple opening network trained with 400 images of flicker data to verify properties, (b) trained structuring elements of erosion layer and dilation layer, respectively.

The experiment is carried out following a 10-fold cross-validation strategy. That means the total image set is divided into 10 groups, and images of 9 groups are used to train the network while the images of the remaining group are used as test images. It is repeated 10 times with each group as the test set and the result of all 400 images are accumulated while those are considered as test images. A sample pair of SEs trained in the network is shown in figures 2.1(b) and 2.1(c).

Verifying 'Increasing' property: Given an image X we generate an intermediate image X' as

$$X'(x,y) = kX(x,y) + \mathcal{N}(0,1)$$

where $k > 0$ and $\mathcal{N}$ is a noise image whose pixel values range from 0 to 1. Finally, X' is clipped between 0 and 1 inclusive to get the final noisy image $\hat{X}$ with bright noise. Thus $X < \hat{X}$, and increasing property ensures that $\gamma(X) \leq \gamma(\hat{X})$. So the opening network fails to satisfy this property if $\gamma(X)(x,y) - \gamma(\hat{X})(x,y) > 0$ for any (x,y). Hence, for u-th image X_u we calculate error $\mathcal{E}_{Iu}$ for increasing property as

$$\mathcal{E}_{Iu} = \frac{1}{|X_u|} \sum_{x,y} \gamma(X_u)(x,y) - \gamma(\hat{X}_u)(x,y) \qquad \text{if } \gamma(X_u)(x,y) - \gamma(\hat{X}_u)(x,y) > 0 \qquad (2.14)$$

where $|X_u|$ denotes the size of image X_u. Histogram of this error $\{\mathcal{E}_{Iu}|u = 1, 2, \cdots 400\}$ is shown in Figure 2.2(a). The mean and standard deviation of $\mathcal{E}_{Iu}$ are 0.000 and 0.000, respectively, which implies that the network satisfies the increasing property.

Verifying 'Idempotent' property: As before an Image X feed to trained opening network produces $\gamma(X)$. Now if we feed $\gamma(X)$ again to the network, we get $\gamma(\gamma(X))$. To hold the idempotent property, $\gamma(X)$ should be equal to $\gamma(\gamma(X))$. In other words, the network fails to satisfy idempotent property if $|\gamma(X)(x,y) - \gamma(\gamma(X))(x,y)| \neq 0$, for any (x,y). Thus, for each u-th image X_u we calculate error $\mathcal{E}_{du}$ as

$$\mathcal{E}_{du} = \frac{1}{|X_u|} \sum_{x,y} |\gamma(X_u)(x,y) - \gamma(\gamma(X_u))(x,y)| \qquad (2.15)$$

(a) Increasing Error   (b) Idempotent Error   (c) Anti-extensive Error
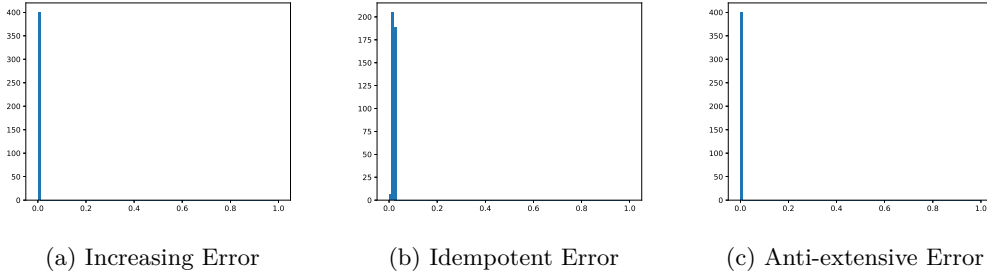
Figure 2.2.: The distribution of (a) Increasing error, (b) Idempotent error, and (c) Anti-extensive.

Histogram of this error $\{\mathcal{E}_{du}|u = 1, 2, \cdots 400\}$ is shown in Figure 2.2(b). Mean and standard deviation of $\mathcal{E}_{du}$ are 0.019 and 0.004, respectively. Since these values are not equal to zero (0) as before, we can at most say that the network very closely satisfies the idempotent property and expect it to satisfy fully if the network is trained better with a larger number of images.

Verifying 'Anti-extensive' property: This property may be verified in exactly similar way for increasing property because anti-extensive property states that $\gamma(X_u)(x, y) \leq X_u(x, y)$. Hence, corresponding error $\mathcal{E}_{au}$ may be computed as

$$\mathcal{E}_{au} = \frac{1}{|X_u|} \sum_{x,y} \gamma(X_u)(x, y) - X_u(x, y) \mid \gamma(X_u)(x, y) - X_u(x, y) > 0 \tag{2.16}$$

Histogram of this error $\{\mathcal{E}_{au}|u = 1, 2, \cdots 400\}$ is shown in Figure 2.2(c). The mean and standard deviation of $\mathcal{E}_{au}$ are 0.000 and 0.000, respectively, which implies that the network satisfies the anti-extensive property.
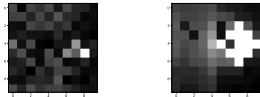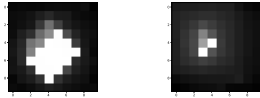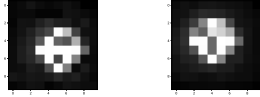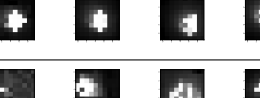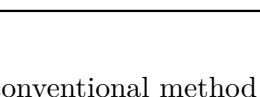
Note that we could have trained the same SE for both erosion and dilation layers to realize opening (and similarly closing) network, but in that case, we need to define the sequence of operations to solve a given problem. On the other hand, training the SEs independently for each layer allows us to develop the required network to solve the problem at hand.

Next, we show how different networks evolve from the elementary morphological layers based on the trained SEs.

Implementing various morphological network using common framework

Here we again take the same set of grayscale images and try to simulate various morphological networks using a sequence of dilation and erosion layers. To generate the ground truth, first, we apply a few morphological operations in a conventional way on the set of images. Then we build a network architecture as a sequence of dilation and erosion layers. We train the morphological network, especially, the structuring elements using the back-propagation

Table 2.1.: Illustrates the efficacy of proposed morphological network. Network architecture given in 2nd column tries to simulate the operation suggested in 1st column. Corresponding learned SEs are shown in 3rd column. (For detail see the text.)

Morphological Operation to be Achieved	Network Architecture	Learned Structuring elements from the first layer
Dilation	$X \rightarrow E^1_{10 \times 10} \rightarrow D^1_{10 \times 10} \rightarrow Y$	
Erosion	$X \rightarrow E^1_{10 \times 10} \rightarrow D^1_{10 \times 10} \rightarrow Y$	
Opening	$X \rightarrow E^1_{10 \times 10} \rightarrow D^1_{10 \times 10} \rightarrow Y$	
Opening	$X \rightarrow E^1_{10 \times 10} \rightarrow D^1_{10 \times 10} \rightarrow E^1_{10 \times 10} \rightarrow D^1_{10 \times 10} \rightarrow Y$	
Closing	$X \rightarrow E^1_{10 \times 10} \rightarrow D^1_{10 \times 10} \rightarrow E^1_{10 \times 10} \rightarrow D^1_{10 \times 10} \rightarrow Y$	

algorithm and mean squared error loss between the output by conventional method and the output produced by the network.

In Table 2.1, we have shown the SEs that are learned based on assigned task and simulated input and output pair. The first column indicates the morphological operations used to generate the output image (Y) from the input image (X). The second column shows the network architecture we have employed to achieve the operation stated in the first column. After training the network (for both the required operation and the associated SE), it is expected to generate the desired output Y given the input X. The learned SEs are shown in the third column. Note that in case of redundant or unnecessary layer, corresponding SE would be an impulse, which realizes an identity transform. For example, in the first row, we try to simulate 'dilation' operation, while the network architecture consists of an erosion layer followed by a dilation layer. After convergence, the trained SE for erosion is found to be approximately an impulse; while SE for the dilation layer is a disk of desired dimension. Similar network is used to simulate erosion and opening as shown in the second and third rows, respectively. The trained SEs justifies their objectives. In the fourth and fifth rows, networks with four layers are used to simulate opening and closing. In the fifth row, it can be seen that the SEs of the first layer (erosion) and the fourth layer (dilation) are approximately impulses as these layers are redundant and unnecessary, and so are expected

to implement identity transform; whereas in the second (dilation) and the third (erosion) layers the learned SEs are similar to disks. However, in the fourth row, combination of first and second layers or combination of third and fourth layers can simulate opening. Secondly, opening is an idempotent operation. So we see trained SEs are more or less similar and are of significant size in all four layers.

In this arrangement We obtain a single dilated or eroded feature map after each layer. However, by using multiple SEs at each layer, multiple feature maps can be generated for performing complex tasks like de-raining and de-hazing.

Though in this simulation experiment, the input and output of the network are both graylevel images, this may be extended to colour image where both input to our network and corresponding output are in (R,G,B) format. We do not apply any strategic treatment to handle colour information, rather let the network to learn the three-dimensional SE of size $A \times B \times 3$ so that it can produce desired output. Second, we have used multiple SEs in each layer. It may help to generate and process features from the colour image and produce desired output. Moreover, in this experiment we use a single SE at each layer, so single feature map is obtained after each layer. By using multiple SEs at each layer, multiple feature maps can be generated for performing complex tasks like de-raining and de-hazing as described in the next sections.

In the next section we have shown the performance operating with same structuring elements in opening closing operations.

2.2.5. Opening closing network with same structuring kernels in dilation and erosion

We know that opening and closing operations are originally defined by using same structuring elements for dilation and erosion operations. However, in this work We have built morphological network to achieve the effect of opening or closing operations using dilation and erosion layers but with separately trained structuring kernels. This allows us more variability in the order of opening and closing operations to create suitable alternating sequential filters for application. We have already shown that the proposed opening and closing networks with separately trained structuring kernels produce results that closely satisfy expected properties of opening and closing. However, it is of interest to see how the opening or closing networks perform with same structuring kernels knowing fully that in that case order of operation has to be decided *apriori* for a specific application.

Here we study morphological opening and closing networks with the same structuring kernels in corresponding dilation and erosion layers. We also compare the performance of this network with that of the networks using separately trained structuring kernels in dilation and erosion layers. To train the network, we have chosen a classical problem,

namely image de-noising considering different configuration of salt and peeper noise, where order of opening and closing is known. As stated we have corrupted the image with different types of noise such as salt noise, pepper noise, and salt-and-pepper noise. Experimental comparison is done between the following two networks: $(ED)_{5\times5} \rightarrow (DE)_{5\times5}$ and $(E_{5\times5}D_{5\times5}) \rightarrow (D_{5\times5}E_{5\times5})$. The former network denotes opening operation followed by closing-operation with same structuring elements, whereas the latter one is generalized opening-closing network proposed by us. In figure 2.2 we have shown the comparison. It can be seen that opening followed by closing network performs better in white noise than black noise removal. In the 3rd row in table 2.2, it can be seen that learned structuring elements opening operation is almost like a spike (unity transform), which allows the input to be propagated to the next layer almost unaffected. The next closing operation is activated because the objective is to remove black noise. In the last two row it can be seen that all the structuring elements for opening closing operations are equally activated due to the presence of both white and black noise in the input.

It can be seen that, though the performance of both the networks are comparable in terms of SSIM, the opening or closing operation with same structuring element performs slightly better in terms of PSNR than the generalized opening closing network for image de-noising. However, in the former case the order of opening and closing should be predefined. It would be non-trivial in case of complex objective. In fig 2.3, we have shown few output of both the opening closing networks for visual comparison. It can be seen that the network is able to retain the detailed texture in the image also removed different kind of noise with both the networks.
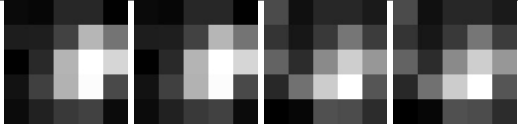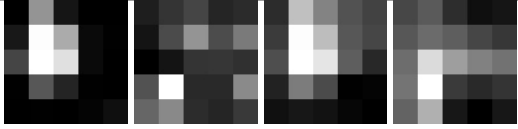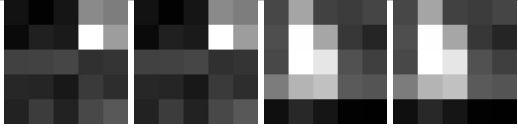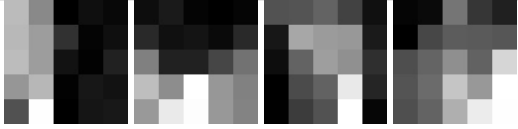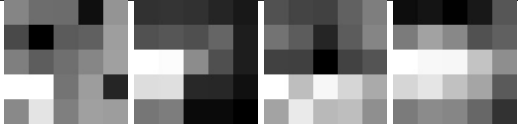
Next, we try to build up generalized opening closing network for image de-raining and image-dehazing.

2.2.6. Image De-raining

The degradation of rainy images depends on several factors such as raindrop size, track of raindrops, rain density and lighting condition. There may be other types of noises present simultaneously in the input image. Morphological filters such as opening and closing are capable of removing noise from an image while preserving the edges. Removing raindrops from an image can be considered as removing mostly bright noise of particular shape and size from the image. This suggests that alternate sequence of dilation and erosion layers, forming opening/closing operations, in the morphological network should be able to realize the de-raining operations by learning the appropriate SEs. However, it is challenging to know the size and shape of the raindrops and other factors beforehand.

Hence, for the purpose, we propose, a morphological network architecture consisting of a sequence of multiple pairs of dilation and erosion layers, which may result in alternating

Table 2.2.: Network architecture given in 1st column. Corresponding learned SEs are shown in 2nd column. $\circ(ED)_{5\times5} \to \bullet(DE)_{5\times5}$ denotes opening and closing operation respectively with structuring element of size 5×5.

Network Architecture	Learned Structuring elements from the first layer	PSNR	SSIM
$\circ(ED)_{5\times5} \to \bullet(DE)_{5\times5}$ (White Noise)		25.44	0.92
$E_{5\times5} \to D_{5\times5} \to D_{5\times5}$ $\to E_{5\times5}$ (White Noise)		24.86	0.92
$\circ(ED)_{5\times5} \to \bullet(DE)_{5\times5}$ (Black Noise)		26.9	0.94
$E_{5\times5} \to D_{5\times5} \to D_{5\times5}$ $\to E_{5\times5}$ (Black Noise)		23.12	0.93
$\circ(ED)_{5\times5} \to \bullet(DE)_{5\times5}$ (White and Black Noise)		24.3	0.89
$E_{5\times5} \to D_{5\times5} \to D_{5\times5}$ $\to E_{5\times5}$ (White and Black Noise)		23.66	0.86

sequential filters (ASF) in parallel paths. Schematic diagram of the network is shown in Fig. 2.4. The output feature maps from the two paths are then linearly combined to get the output maps. This step is essential for recovering undesired removal of some features by opening or closing filters. Finally, sigmoid activation function is applied to it to produce the final output (image). In Table 2.3 we have shown three architectures, which we have used for our experiments. In [53], the authors have shown that opening network, as expected, is more effective in removing bright noise. Usually closing network removes dark noise, which does not have much to do in this application. However, to make our network more general, we take a linear combination of the output of opening path and closing path and make a single generic architecture. Since the output of the de-rained image should be of the same size as the input image, appropriate zero padding is used in each layer. For the color image processing, a number of feature maps after each dilation layer and erosion layer using as many SEs. Multiple feature maps help in propagating the color information to the output.

(a) Input   (b) Generalized open-close net   (c) Same SE open-close net

Figure 2.3.: Comparison of output with same structuring element opening-closing network and the generalized opening-closing net. First row, middle row and last row contains only salt, pepper and salt-and-pepper noise respectively.

During training, we initialize all the SEs and the weights of the linear combination layers with random numbers. To train the network, we define a loss using the structural similarity index measure (SSIM) between predicted output P and the ground truth T of the data defined as

$$\mathrm{SSIM}(P,T) = \frac{(2\mu_P \mu_T + c_1)(2\sigma_{PT} + c_2)}{(\mu_P^2 + \mu_T^2 + c_1)(\sigma_P^2 + \sigma_T^2 + c_2)} \tag{2.17}$$

Here μ_P and μ_T are the mean for image P and T, respectively; while σ_P and σ_T are their standard deviations. The term σ_{PT} is the co-variance between the images, and c_1 and c_2 are constants, whose values are set to 0.0001 and 0.0009, respectively to avoid divide by zero. We use the structural dissimilarity (DSSIM) as the loss function considering all small

Figure 2.4.: The network for image de-raining. It consists of 2 parallel paths containing a complementary sequence of Erosion and Dilation layers. The structuring element for each operation is 8×8. The outputs from each of the paths are taken together for linear combination to produce three feature maps, which pass through *sigmoid* activation to get the final output.

Table 2.3.: The architectures of paths shown in Fig. 2.4. $D_{8 \times 8}^{8}$ denotes a layer with 8 dilation operators with separately trainable SEs of size 8×8. $E_{8 \times 8}^{8}$ is defined similarly. In the last opening layer of closing network and last closing layer of opening network, we have taken c number of structuring elements where c is 1 for grayscale image and 3 for color image processing task.

Description of the MorphoN	
Closing Network	$X \to D_{8 \times 8}^{8} \to E_{8 \times 8}^{8} \to D_{8 \times 8}^{8} \to E_{8 \times 8}^{8} \to D_{8 \times 8}^{8} \to E_{8 \times 8}^{c} \to Y$
Opening Network	$X \to E_{8 \times 8}^{8} \to D_{8 \times 8}^{8} \to E_{8 \times 8}^{8} \to D_{8 \times 8}^{8} \to E_{8 \times 8}^{8} \to D_{8 \times 8}^{c} \to Y$
Opening-Closing Network	

patches of the output and those of corresponding ground truth for training the network. So the loss function is defined as

$$Loss_{total} = \text{DSSIM}(I_{out}, I_{gt}) = \frac{1}{M} \sum_{i} \frac{1 - \text{SSIM}(P_{out}^{i}, P_{gt}^{i})}{2} \tag{2.18}$$

where I_{out} and I_{gt} are the output image and the ground truth image, and P_{out}^{i} and P_{gt}^{i} are the ith patches, respectively. There are a total of M patches. In our experiment, we have taken the patch size as 10×10. We minimize the loss function $Loss_{total}$ by back-propagation algorithm and learn the parameters of the network.

Figure 2.5.: The network for image de-hazing. It consists of 2 parallel paths containing a complementary sequence of Erosion and Dilation layers. The structuring element for each operation is 4×4. The output from each of the paths is concatenated and passed through three 1×1 convolution layers. The airlight and transmittance map is generated by sending the output to 2 parallel sub-networks. The output from each of the sub-network is linearly combined to get the final output.

2.2.7. Image De-Hazing

When light rays travel through a turbid medium, it undergoes a phenomenon known as scattering, wherein the light is scattered in different directions due to its interactions with particles, such as dust and aerosols floating in it. Haze occurs when the concentration of these particulate matters exceed a certain threshold. Any image taken in hazy conditions suffers from visibility degradation such as reduced contrast, saturation attenuation and color shifting. Besides, scattered environmental light appears like a veil over the scene. These days, image de-hazing has become one of the trending problems. Here we intend to restore such hazy images using the opening-closing network. The observed hazy image can be physically modelled using the following equation [38]

$$I(x) = t(x)J(x) + (1 - t(x))A, \qquad (2.19)$$

where $I(x)$ is the observed or recorded intensity of the hazy image at location x and $J(x)$ is the intensity of corresponding non-hazy (ideal) image at location x. A is the airlight, which characterizes the constant environmental illumination. $t(x)$ is the transmittance coefficient that determines the amount of light reaching the observer (camera) from the objects after travelling through the medium. Transmittance $t(x)$ intuitively measures the amount of haze present in a particular location x and, in general, depends upon the depth

of the scene. We modify equation 2.19 as

$$I(x) = t(x)J(x) + K(x), \tag{2.20}$$

where the bias like term $K(x)$ is the space-variant airlight representing $(1 - t(x))A(x)$. Also note that, since $0 \leq t(x) \leq 1$, contrast of $J(x)$ is reduced to $J(x)t(x)$. It is known that morphological opening can estimate local bias if flat SE of appropriate size is used. Second, Opening (resp. Closing) operator can detect local bright (resp. dark) features whose scaling affects the contrast. The morphological network is employed to estimate the airlight map as well as the transmittance map in order to recover the haze-free image. As shown in Fig. 2.5, we have taken two separate paths: one for opening and the other closing. The outputs of the paths undergo linear combination. The network produces the transmittance map $t(x)$ and the airlight map $K(x)$ as output. Since $0 \leq t(x) \leq 1$ and $0 \leq K(x) \leq 1$, in the last layer of transmittance path and also of airlight path we have employed sigmoid activation function to limit the estimated values of $t(x)$ and $K(x)$ within the said range. Once $t(x)$ and $K(x)$ are estimated, we can determine de-hazed image as

$$J_{out}(x) = \min\left(\frac{I(x) - K(x)}{t(x)}, 1\right) \tag{2.21}$$

During training of the network, with given $t(x)$ and $K(x)$, hazy image is generated as

$$I(x) = J(x)t(x) + K(x); \tag{2.22}$$

Now, given a pair of hazy and haze-free clear images, the network learns the SEs and also the weights of the linear combination layer. Network gives $\hat{t}(x)$ and $\hat{K}(x)$ at every iteration of the training process. We reconstruct the hazy image $\hat{I}_{out}$ and estimated de-hazed image $\hat{J}_{out}$ based on these $\hat{t}(x)$ and $\hat{K}(x)$ using the equations 2.21 and 2.22. We define the loss function L, which is very similar to bi-directional consistency loss [55] given by

$$L = L_1 + L_2, \tag{2.23}$$

where,

$$L_1(x) = \text{DSSIM}(I, \hat{I}_{out}), \tag{2.24}$$
$$L_2(x) = \text{DSSIM}(J, \hat{J}_{out}); \tag{2.25}$$

DSSIM is calculated using equation 2.18. We minimize the loss L and learn the network parameters. In the next section, we present experimental results to justify our claim.

2.3. Experimental results

In this section, we evaluate performance of the proposed network both quantitatively and qualitatively on de-raining and de-hazing tasks.

2.3.1. Experimental Setup and Data Set

We have implemented the network in Python using Keras with TensorFlow library at the backend. We carried out our experiments on Intel machine with a GPU of memory 12GB. For all the experiments, we have initialized the structuring elements randomly using standard glorot uniform initializer [27]. To minimize the loss, we have used RMSProp optimizer for all the networks.

Training of image de-raining network is done with benchmark rain dataset [24]. This rain dataset has $1,000$ clean images. For each clean image, there are 14 different synthesized rainy images with different streak orientations and sizes. So, a total of 14000 sample images are available. We have considered 80% of the data for the training and 10% of the data for validation, and the rest 10% is used as test data. Images in the dataset are of different sizes. However, as the network takes fixed-size input, all the images are resized to 416×416 through bi-linear interpolation.

For de-hazing, we have used O-HAZE [3] dataset for training. This dataset was first proposed in O-HAZE [4] challenge. The dataset has 35 images for the training and five images for validation and five images for testing. In this set, images are of different sizes, and the sizes are of the order of 5500×3500 pixels. So we have resized each image to 1024×1024. However, while testing, we have taken the full resolution of the images. Since the O-HAZE dataset is very small, we have also used NYU dataset of D-Hazy dataset [5] for the training. For quantitative evaluation of our Opening/Closing network, we have used validation dataset of the O-HAZE as test set because their ground truths are available. We also test our network in Fattal and Middlebury dataset of D-Hazy [5].

We have trained both the de-rain and de-haze network until the loss is converged. Note that, the loss function involving max/min (due to dilation/erosion) operation is only piece-wise differentiable. In practice, we can find the sub-gradient of each morphological operation and back-propagate through the network. In the next two subsections, we have evaluated our Morphological network developed for De-hazing and De-raining network.

2.3.2. Results of Image De-Raining

Here we present the quantitative and qualitative evaluation of morphological network for image de-raining.

Table 2.4.: Test results of de-rain network, quantified by SSIM and PSNR, on both grayscale and color images of Rain dataset. **Closing Net** and **Opening Net** show the results of the two parallel paths when trained individually. The result of de-rain network (**opening-closing net**) is compared with **CNN**

	Metric	Base Line	Closing Net	Opening Net	Opening-Closing Net	CNN	MorphoN[53]
Gray	SSIM	0.85	0.90	0.90	0.91	**0.93**	0.92
	PSNR	24.41	26.00	25.99	27.29	**29.24**	28.03
Color	SSIM	0.84	0.88	0.87	0.89	**0.91**	-
	PSNR	24.06	24.81	24.28	25.01	**27.33**	-

Quantitative Evaluation

In order to evaluate quantitatively the opening-closing network applied on rainy images, we have used two objective measures such as SSIM [95] and Peak signal-to-noise ratio (PSNR). In [53], image de-raining is done only on grayscale images. We have extended that work for de-raining of color images. The extension is not straightforward and a minor modification of network is needed to preserve hue in the image. We trained the network on the rainy image dataset until the convergence of training loss is achieved. Opening Net, Closing Net, and the opening-closing net are trained separately to study their relative performances. The architectures of these networks are illustrated in Table 2.3. The estimated de-rained image from each network is quantitatively evaluated against the rain-free image available as ground truth in the dataset. The average values of SSIM and PSNR for grayscale and color test images, are reported in Table 2.4. In [53], we have shown that opening path (network) performs better than the closing path (network) as the opening filter removes bright noise. However, as shown in Table 2.4 closing network performs similarly well as opening network. We believe this is because we have taken multiple dilation/erosion layers and problem dependent learning of SEs has created effectively an opening path emphasizing bright noise removal. We have also compared our results with a standard convolutional neural network (CNN) of U-net architecture [74]. This Table also reveals that the opening-closing network gives results comparable to that of CNN with U-net architecture. However, we observe that CNN has more generalization capability than the opening-closing network, but needs to train a huge number of parameters, which incurs a very high cost. It can also be seen that in the PSNR metric, we do not get much improvement for both grayscale and color images. It may be because of training the network with SSIM loss only. Incorporating mean squared error in the loss may result in better PSNR.

Qualitative Evaluation

For qualitative evaluation, we have shown the results of different networks applied on a large number of images to nine evaluators (research fellows in our lab) and asked them to

(a) Input Frame   (b) Closing Net   (c) Opening Net   (d) Opening-Closing Net   (e) CNN   (f) Ground Truth
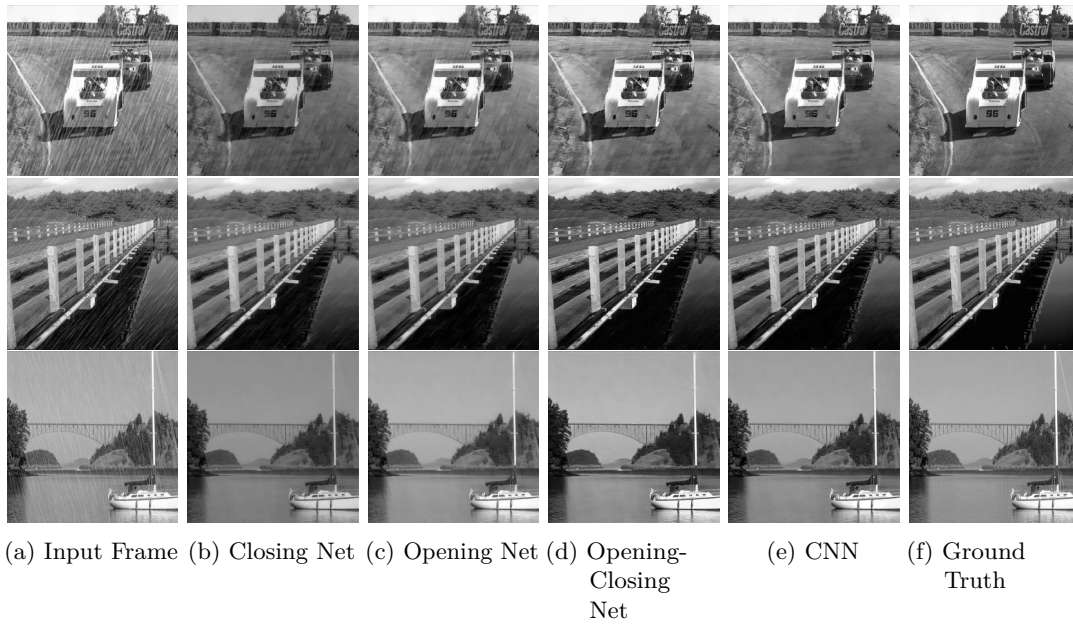
Figure 2.6.: Result of the De-rain network over grayscale images from Rain dataset. The output from Closing Net, Opening Net, De-rain network and CNN, along with input image and ground truth are shown for qualitative comparison.

(a) Input Frame   (b) Closing Net   (c) Opening Net   (d) Opening-Closing Net   (e) CNN   (f) Ground Truth
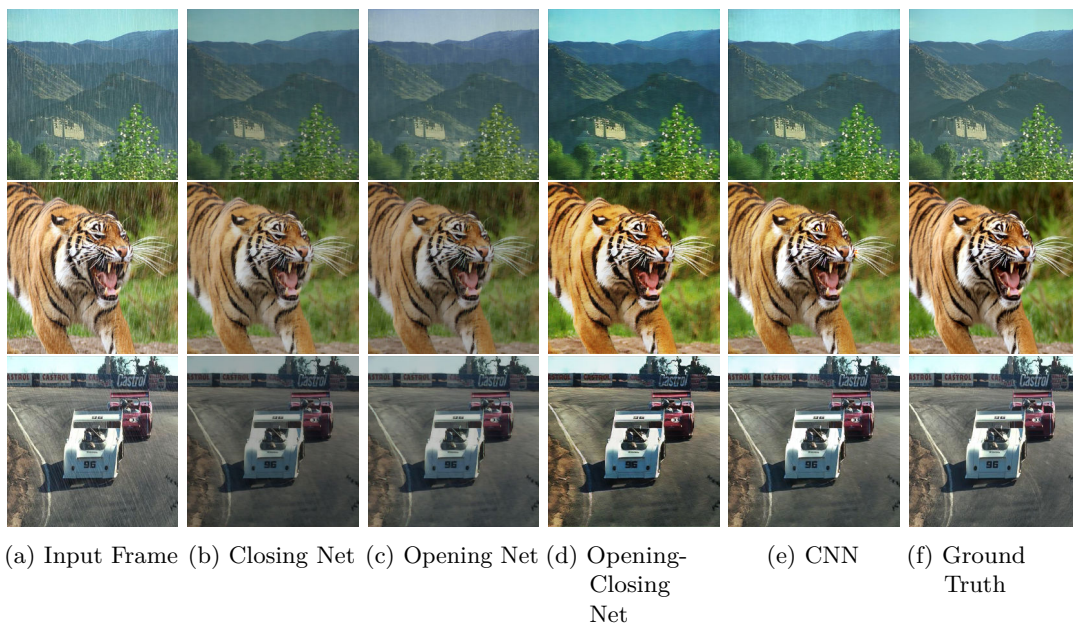
Figure 2.7.: Result of De-rain network over color images from Rain dataset. The output from Closing Net, Opening Net, De-rain network and CNN, along with input image and ground truth are shown for qualitative comparison.

| (a) Input Frame | (b) Closing Net | (c) Opening Net | (d) Opening-Closing Net | (e) CNN |
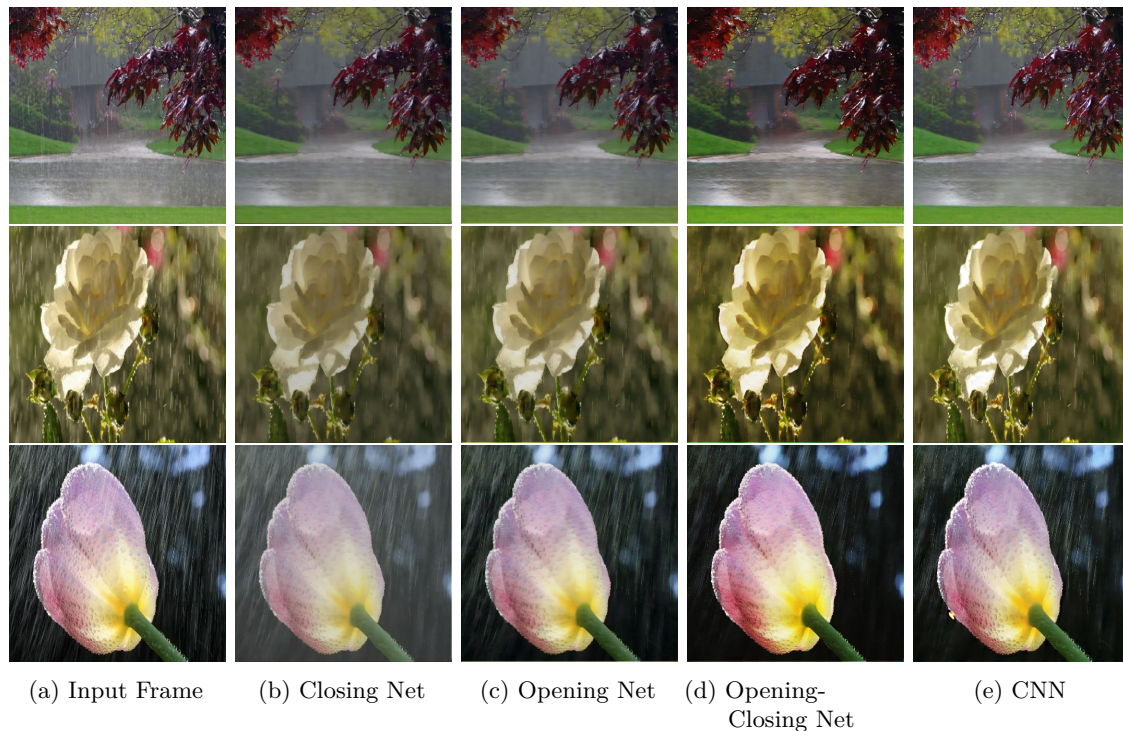
Figure 2.8.: Results from Closing Net, Opening Net, De-rain network and CNN on real rainy images

assign ranks to the output. Order of average rank is similar to quantitative evaluation. That is CNN comes out to be the best closely followed by opening-closing network. Results of opening network and closing network are comparable. Here we display results for only 3 grayscale images in Fig. 2.4 and 3 color images in Fig. 2.7. In both figures, we compare the output of each network with ground truth.

In grayscale images (Fig. 2.4) it may be noted that though the opening-closing network did remove raindrops successfully, blobs of uniform intensity are present. This effect is a bit more severe in case of opening network and closing network (especially in the top row images). In the case of color images, as seen in Fig. 2.7, the opening-closing network can effectively de-rain and also reproduces the vivid colors with desired contrast. The output of the closing network has less contrast than the output of the opening network, whereas the opening-closing Net is able to address the issue. Secondly, the blob effect, as seen in grayscale images, are not visible here.

The results shown above are on images from benchmark rain dataset [24], where rain images are synthesized from clean images; hence groundtruth for images are available. We also show few results on real rain images in Fig. 2.8 for which the groundtruth are not available. The opening-closing network is able to clear the rain perfectly, whereas the output of CNN(U-Net) has surprisingly few small rain streaks in some images (see bottom

row of the figure). This may be due to presence of large number of parameters in CNN (U-net) leading to over-fitting.

Table 2.5.: Quantitative evaluation on validation dataset of O-HAZE images in terms of PSNR and SSIM metrics.

Method	36.png PSNR/SSIM	37.png PSNR/SSIM	38.png PSNR/SSIM	39.png PSNR/SSIM	40.png PSNR/SSIM	Average PSNR/SSIM
DCP[30]	18.18/0.45	16.09/0.49	14.12/0.08	12.88/0.36	14.21/0.39	15.10/0.35
CAP[107]	17.47/0.50	16.17/0.45	15.14/0.18	14.80/0.41	16.37/0.57	15.99/0.42
MS-CNN[70]	16.59/0.49	15.76/0.43	13.25/0.19	12.79/0.40	16.53/0.56	14.99/0.41
DehazeNet[13]	16.92/0.43	14.98/0.48	15.54/0.34	17.65/0.48	17.04/0.54	16.43/0.45
AOD-Net[42]	17.10/0.45	16.47/0.39	16.12/0.12	15.04/0.34	15.95/0.50	16.13/0.36
DCPDN[102]	17.14/0.44	15.29/0.42	14.66/0.11	15.24/0.36	17.78/0.52	16.02/0.36
Zhang et al.[103]	**24.67**/0.73	**22.41**/0.66	**23.75**/0.72	**21.91**/0.63	**22.29**/0.68	**23.00**/0.68
Ours	20.22/**0.75**	21.11/**0.75**	19.45/**0.75**	19.63/**0.76**	14.57/0.64	19.00/**0.73**

Table 2.6.: Quantitative evaluation of opening-closing network with state of art algorithms on 12 images of Fattal dataset in terms of SSIM and PSNR metrics.

Image	[29] PSNR/SSIM	[20] PSNR/SSIM	[69] PSNR/SSIM	[10] PSNR/SSIM	[41] PSNR/SSIM	Opening-Closing Net PSNR/SSIM
church	11.16/0.78	**21.43/0.96**	14.17/0.87	15.68/0.89	9.44/0.62	12.32/0.82
couch	18.4/0.86	**20.8/ 0.9**	17.99/0.88	17.24/0.87	16.77/0.83	16.04/0.81
dolls	19.73/ 0.85	**21.29**/0.77	16.93/**0.86**	15.69/0.83	17.21/0.85	17.76/0.81
flower1	14.1/0.88	**30.01/0.98**	9.08/0.43	12.15/0.72	12.22/0.79	14.71/0.88
flower2	14.37/0.86	**31.94/0.99**	10.81/0.6	11.86/0.68	13.13/0.79	15.16/0.85
lawn1	13.84/0.8	**24.49/0.97**	14.37/0.83	14.78/0.86	11.32/0.69	11.63/0.77
lawn2	11.2/0.74	**24.94/0.97**	13.29/0.77	15.32/0.87	10.98/0.68	11.55/0.77
mansion	17.45/0.87	**26.96/0.97**	17.69/0.89	17.33/0.87	14.24/0.7	15.73/0.86
moebius	12.66/0.78	**19.01/ 0.9**	16.36/ 0.9	14.58/0.85	13.22/0.77	13.82/0.86
raindeer	18.12/0.83	**26.22/0.94**	16.82/0.81	16.59/0.82	16.53/0.8	16.41/0.85
road1	12.95/0.8	**25.74/0.96**	14.11/0.84	16.3/0.88	11.75/0.66	12.05/0.78
road2	15.84/0.84	**23.6/0.96**	16.45/0.88	18.22/0.9	11.96/0.62	12.44/0.78
Average	14.98/0.82	**24.7/0.94**	14.84/0.8	15.48/0.84	13.23/0.73	14.13/0.82

2.3.3. Results of Image De-Hazing

In this subsection, we evaluate the performance of the opening-closing network quantitatively and qualitatively with respect to the de-hazing problem. De-hazing problem is relatively more wi attempted. So there exist quite a few benchmark datasets and many good algorithms. We have compared our method with multiple classical methods like Dark channel Prior(DCP) [30],color attenuation prior(CAP) [107] and multiple deep learning methods like Multiscale CNN(MS-CNN) [70], DehazeNet [13] AOD-Net [42] and DCPDN [102]. DCP method is based on the observation that in hazy images intensity of color channels is uniformly contributed by the air-light. In CAP method, they have approximated the haze by the difference of brightness and saturation. Recent deep learning based methods like MS-CNN and DehazeNet predict transmittance map using the network. AOD-Net

Table 2.7.: Quantitative results of opening-closing network on Middlebury portion of D-Hazy dataset with PSNR and SSIM metrics. The opening-closing network has the second highest average PSNR and SSIM values over the other methods

Image	[29] PSNR/SSIM	[69] PSNR/SSIM	[10] PSNR/SSIM	[41] PSNR/SSIM	Opening-Closing Net PSNR/SSIM
Adirondack	16.02/0.82	14.39/**0.89**	**16.74**/0.88	14.18/**0.89**	16.54/0.87
Backpack	14.4/0.85	**16.21**/0.87	12.24/0.82	16.1/**0.91**	14.22/0.85
Bicycle1	12.39/0.81	20.66/0.93	12.61/0.82	**23.21**/**0.96**	14.2/0.86
Cable	**12.95**/**0.7**	7.65/0.64	9.93/0.63	6.95/0.64	12.54/0.68
Classroom1	20.17/0.87	10.91/0.74	**20.95**/**0.89**	10.02/0.72	12.34/0.76
Couch	**18.68**/**0.81**	10.13/0.61	13.76/0.7	10.56/0.63	12.01/0.64
Flowers	**17.73**/0.89	10.47/0.78	17.45/**0.9**	9.25/0.76	14.85/0.83
Jadeplant	**13.48**/**0.69**	7.78/0.6	7.06/0.65	7.65/0.59	13.1/0.64
Mask	15.88/0.89	14.15/0.85	14.18/0.84	14.3/**0.91**	**15.88**/0.85
Motorcycle	13.81/0.79	13.2/0.81	11.6/0.62	12.25/**0.82**	**14.49**/0.79
Piano	**18.66**/**0.86**	12.4/0.71	15.08/0.78	13.89/0.75	16.8/0.79
Pipes	**15.52**/**0.79**	10.9/0.68	13.81/0.74	10.34/0.69	15.42/0.78
Playroom	**17.7**/**0.85**	13.42/0.77	17.64/ 0.83	13.24/0.78	15.47/0.79
Playtable	**18.58**/**0.9**	15.09/0.86	16.63/0.88	14.73/0.86	15.99/0.86
Recycle	12.5/0.82	**18.3**/**0.95**	13.43/0.88	16.62/0.9	13.03/0.87
Shelves	15.47/0.83	**20.43**/**0.94**	16.9/0.88	16.52 ?/0.92	16.88/0.87
Shopvac	**13.87**/**0.8**	7.62/0.66	11.58/ 0.78	6.89/0.64	11.15/0.73
Sticks	16.96/0.9	**20.5**/**0.96**	20.41/0.93	19.13/**0.96**	19.28/0.93
Storage	17.38/**0.88**	11.23/0.82	16.36/**0.88**	10.24/0.79	**17.71**/0.87
Sword1	15.06/0.87	**15.48**/**0.91**	12.57/0.83	14.29/**0.91**	14.37/0.86
Sword2	15.66/0.89	12.89/0.88	14.89/0.88	12.8/ **0.9**	**15.91**/0.89
Umbrella	10.4/0.8	**14.92**/0.9	9.63/0.72	14.58/**0.91**	12.11/0.84
Vintage	14.63/0.86	**19.27**/**0.96**	14.09/0.83	16.82/0.94	13.76/0.85
Average	**15.56**/**0.83**	13.82/0.81	14.33/0.81	13.35/0.82	14.70/0.82

reformulates the haze equation using a single parameter and estimates it using CNN based network. DCPDN algorithm incorporates the concept of image pyramid inside the network.

Quantitative Evaluation

In De-hazing, we quantitatively compare the performance of our opening-closing network with other state-of-the-art methods. We have trained the opening-closing network using the O-HAZE training dataset and tested the network on O-HAZE validation dataset. The result is reported in Table 2.5. The PSNR and SSIM values of the results of different algorithms for each of the 5 images of validation dataset are shown in the table. It is seen that the output of the opening-closing network gives the best results in terms of SSIM in most of the images. However, in terms of PSNR, it ranked 2nd best in the table next to that of [20].

We have also trained the network using NYU portion of D-Hazy dataset and tested it on fattal [20] and Middlebury portion of D-Hazy dataset. We report these results in Table 2.6 and Table 2.7, respectively. From Table 2.6 it is revealed that the performance of

| (a) Input Image | (b) Transmittance | (c) Airlight | (d) Our Output | (e) Ground Truth |
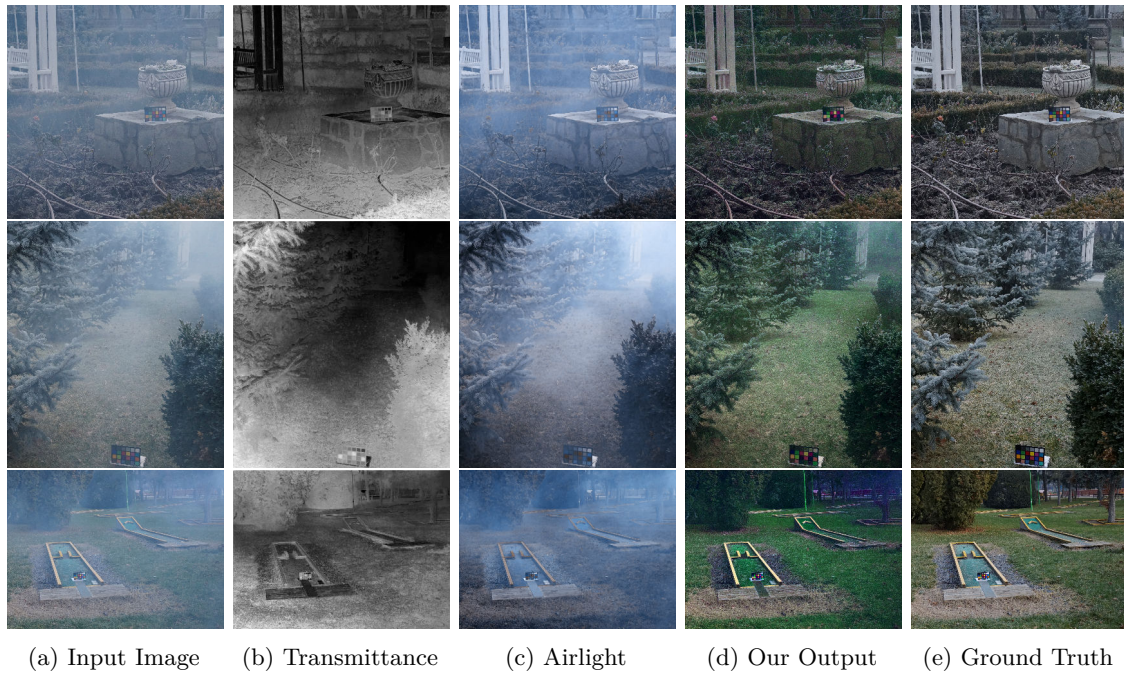
Figure 2.9.: Results of opening-closing de-haze network over O-HAZE dataset. The transmittance and airlight map along with ground truth is shown for qualitative evaluation.

| (a) Input Image | (b) Transmittance | (c) Airlight | (d) Our Output |
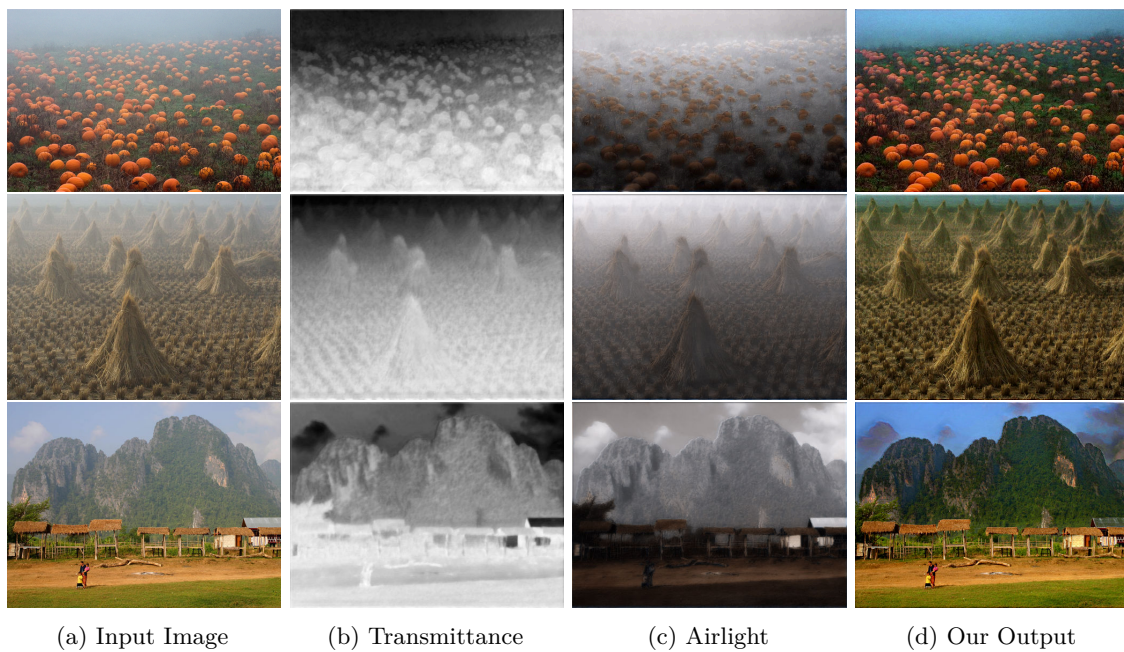
Figure 2.10.: Results of opening-closing De-haze network on real outdoor images. Transmittance and airlight map along with ground truth is shown for qualitative evaluation.

the opening-closing network on Fattal dataset is not as good as that on O-HAZE dataset (Table 2.5). Probably it happens because NYU portion of D-Hazy dataset has only white

airlight, while in Fattal dataset the airlight may not be white. However, in Table 2.7, We see that the performance of our morphological network is comparable to that of state-of-the-art algorithms on Middlebury portion of D-Hazy dataset.

Qualitative Evaluation

For qualitative evaluation of the de-hazed image, we have used the same strategy as has been used for de-rained images. However, comparing the quality of de-hazed images is so difficult that grading is often inconsistent. We show the output of the opening-closing network on only three challenging images from O-HAZE dataset along with groundtruth. The result of our network (fourth column) along with input image (first column) and the groundtruth (fifth column) are shown in Fig. 2.9. The second and third columns show

(a) Input Image   (b) Transmittance   (c) Airlight   (d) Our Output
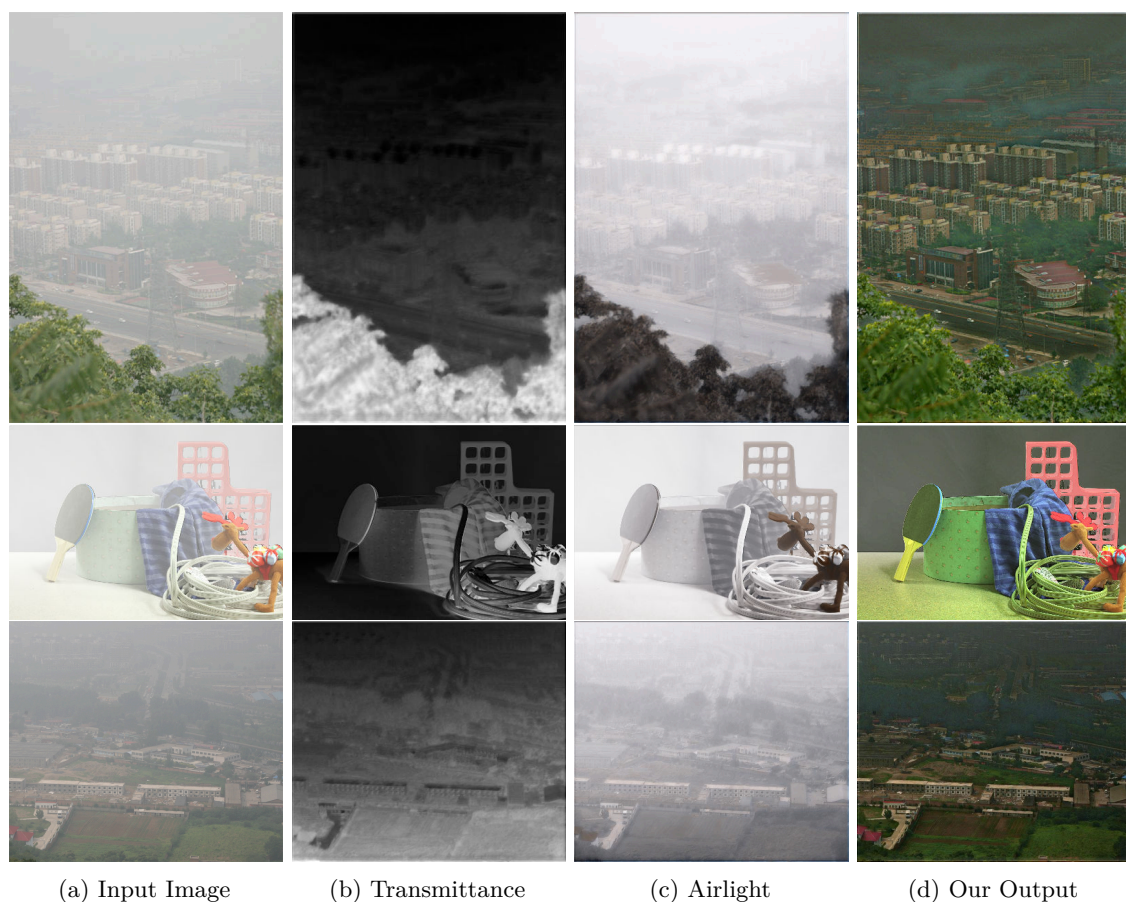
Figure 2.11.: In this figure we have shown few failure cases of opening-closing de-haze network. The predicted transmittance (b) and airlight (c) maps are shown along with output (d). The network is able to remove haze considerably but fails to preserve the color in the image(2nd row). Whereas in 1st and 3rd it removes the white haze but at the same time it makes the image darker.

estimated transmittance map and airlight, respectively. The average subjective quality over results of our method on this dataset given by the evaluators is 'reasonably good' and tends to be of 'low brightness'. It should be noted that most of the state-of-the-art methods assume one of the two parameters (i.e., transmittance and airlight) is either known or can be determined heuristically, and the other (commonly transmittance) is estimated. On the other hand, the opening-closing network estimates both the parameters simultaneously from the input image and de-haze the same by applying the appropriate transformations.

Apart from the known benchmark datasets, we have also applied our morphological network on some real-world hazy images popularly used by the research community. Results on three images are shown in Fig. 2.10. We observe that our network is able to effectively de-haze the images and manages to reproduce the colors faithfully. However, in some instances, it was unable to reproduce lighter shades of some color. This can be attributed to the fact that we haven't used a highly optimized network for de-hazing images. In the real scene, the amount of haze is increased when depth is increased. The depth effect can be clearly seen in the transmittance map (second column (b)). Although the opening-closing network is able to de-haze the image, sometimes it fails and cannot obtain the haze-free image correctly. In Fig. 2.11 we have reported a few failure cases of opening-closing network. In the first row and last row in Fig. 2.11, it can be seen that the output image is darker. This may be because of the overestimation of the transmittance map. In the 2nd row, the estimated haze-free image has its table color green from white. From the estimated airlight and transmittance map, we can see that it considers the white table as haze.

2.4. Summary

In this Chapter, we have renewed the concept of learning morphological structuring elements (SEs) with opening and closing operations. We have built a network that creates the effect of opening and closing with dilation and erosion layers placed sequentially in two different parallel paths and then by taking a linear combination of their outputs. To establish the efficacy of the proposed network, we have exploited it on two classes of image restoration problems, namely de-raining and de-hazing. Like CNN taining methods, the proposed network learns SEs and various weights as parameters that lead to desired morphological network in terms of both operations and SEs. Once learnt, the network then can estimate degrading parameters from the test input image and, subsequently, restore it. Although we get comparable results with state-of-the-art methods, we believe improving loss function may give better results. Second, to minimize the loss function, the structuring elements are trained by back-propagation algorithm. Derivative in back-propoagation is computed by exploiting subgradient method as dilation/erosion operators are defined in terms of piece-wise differentiable max/min operators. Although we have used opening and closing

operations for image restoration task, what should be the optimal network to perform all classification tasks using dilation and erosion operations is still unknown. So, in the next chapter, we will build dense network using dilation and erosion operation and do rigorous theoretical analysis using min or max operation for feature classification.

Part III.

Dense Morphological Network

Chapter 3.

Dense Morphological Network

3.1. Introduction

In the previous chapter, Image restoration is done using opening-closing network. The network is basically a combination of dilation and erosion operations over the image or image features, where dilation and erosion operations are defined in terms of min and max operation, respectively. Most real-world problems are complex in nature and potentially require a combination of elementary transformations. So, a sequence of operations are often employed with appropriate structuring elements. But the choice of the operations and the associated structuring elements require expert knowledge along with some trial-and-error. In this chapter, we work with min and max operations and make a rigorous theoretical analysis of it in one dimensional network form. We call this novel morphological network as *Morph-Net* where 1D dilation and 1D erosion operators are used as neurons.

Unlike neural network, i.e, instead of sum of product, *Morph-Net* compute max (or, min) over the sum (resp. difference). More complex morphological operators are constructed by the composition of these two basic operators. However, to achieve acceptable performance, this proposed morphological network should be generic enough. We have shown here that the network with 1-D dilation and erosion operators, when built in a certain way, is also generic and can realize piece-wise linear function.

We start with the most basic version of a 1D morphological network, where the morphological operators work over the whole input, not locally, to make theoretical analysis tractable. This single layer version of the proposed morphological network contains a layer of both dilation and erosion operations, followed by a layer computing linear combination of the output of these operators. We show that this simple network can represent a piece-wise linear function. Moreover, this network can be trained by usual gradient descent approach. Then we move on to deeper architectures and proved the universal approximation of the network. We have also experimented with the 'soft' version of the max and min operator to alleviate the potential problems of non-differentiality. To experimentally validate the efficacy of this network in real-life applications, we have evaluated its performance on

standard classification datasets like MNIST, Fashion-MNIST, SVHN and CIFAR-10.

Araújo [1] utilized network architecture similar to morphological perceptrons with competitive learning to forecast stock markets. The *argmax* operator was replaced with a linear activation function so that the network could regress forecasts and allowed the use of gradient descent for training. Recently few works [2, 23, 59] have been done in this direction. Nogueira et al. [59] have introduced morphological network with multiple morphological operators for image classification task. Islam et al. [2] used morphological hit-or-miss transform to build a network for similar purpose. For morphological neurons with dendritic structure Zamora and Sossa [100] replaced the *argmax* operator with a softmax function. This overcomes the problem of gradient computation and, therefore, gradient descent optimizer could be employed.

Rest of this chapter is organized as follows. In section 3.2, we have described the most basic form of our proposed morphological network along with the analysis of its capabilities. Then a more complex architecture, namely Deep Morph-net, is presented in Section 3.2.8. Section 3.3 provides the experimental validation of the capabilities of the proposed networks.

3.2. Morphological Network

In this section we introduce the Morphological network or 'Morph-Net', in short. It is a simple feed forward network that consists of dilation and erosion neurons. We begin with defining dilation and erosion neurons, and then describe the simplest version of the network and its capabilities.

3.2.1. Dilation and Erosion neurons

Dilation and Erosion neurons are the building blocks of our proposed network. Given an input $\boldsymbol{x} \in \mathbb{R}^d$ and a structuring element $\boldsymbol{s} \in \mathbb{R}^d$, the operation of **dilation** ($\oplus$) and **erosion** ($\ominus$) neurons are defined, respectively, as

$$\boldsymbol{x} \oplus \boldsymbol{s} = \max_k \{x_k + s_k\} \tag{3.1}$$

$$\boldsymbol{x} \ominus \boldsymbol{s} = \min_k \{x_k - s_k\} \tag{3.2}$$

where x_k denotes k^{th} element of input vector $\boldsymbol{x}$. After computing dilation and erosion we may set a *limiter* or *bias*, say, s_{d+1} to compute final output from dilation and erosion neurons by $\max\{\boldsymbol{x} \oplus \boldsymbol{s}, s_{d+1}\}$ and $\min\{\boldsymbol{x} \ominus \boldsymbol{s}, -s_{d+1}\}$ respectively. Note that this ensures s_{d+1} to be lower bound of the output of dilation neuron, whereas it is upper bound for the erosion; hence, the term 'limiter'. Alternatively we can write it as follows. Let 0 is

appended to the input $\boldsymbol{x}$, i.e, $\boldsymbol{x}' = [\boldsymbol{x}, 0]^T$ and s_{d+1} is appended to $\boldsymbol{s}$, then

$$\max\{\boldsymbol{x} \oplus \boldsymbol{s}, s_{d+1}\} = \max\{\max_k\{x_k + s_k\}, s_{d+1}\} \tag{3.3}$$

$$= \max_{k=1,\dots,d+1}\{x'_k + s'_k\} \tag{3.4}$$

$$= \boldsymbol{x}' \oplus \boldsymbol{s}' \tag{3.5}$$

Where s'_k is a element of structuring element $\boldsymbol{s}'$. Similarly we can get $\boldsymbol{x}' \ominus \boldsymbol{s}'$. It may be argued that $d + 1^{th}$ component is selected if the input has no effect on the output or the function. Note that, the erosion operation may also be written as $\boldsymbol{x}' \ominus \boldsymbol{s}' = -\max_k\{s'_k - x'_k\}$. In these neurons the structuring element ($\boldsymbol{s}'$) is learned in the training phase.

The max and min operators used in the dilation and erosion neurons are only piece-wise differentiable. As a result, only a single element of structuring element is updated at each iteration of backpropagation algorithm making the learning very slow. To overcome this problem we propose to use the soft version of max and min [16] to define soft dilation and erosion operation as follows.

$$\boldsymbol{x}' \hat{\oplus} \boldsymbol{s}' = \frac{1}{\beta} \log \left(\sum_{k+1} e^{(x'_k + s'_k)\beta} \right), \tag{3.6}$$

$$\boldsymbol{x}' \hat{\ominus} \boldsymbol{s}' = -\frac{1}{\beta} \log \left(\sum_{k+1} e^{(s'_k - x'_k)\beta} \right), \tag{3.7}$$

where $\hat{\oplus}$ and $\hat{\ominus}$ denote the soft dilation and soft erosion, respectively, and β is the "hardness" of the soft operations. The soft version can be made close to their "hard" counterpart by making β large enough [16]. Henceforth, for notational convenience we use $\boldsymbol{x}$ and $\boldsymbol{s}$ to denote input and structuring element respectively for dilation (or erosion) with limiter. In other words, dilation-erosion layer includes the limiter.

3.2.2. The morphological block

The simplest form of Morph-Net is the morphological block. It consists of a layer with dilation and erosion neurons (including limiter) followed by a linear combination (Figure 3.1) of their outputs. We call the layer of dilation and erosion neurons as the **dilation-erosion layer** and the following layer as the **linear combination layer**. Let us consider a network with n dilation neurons and m erosion neurons in the dilation-erosion layer followed by c neurons in the linear combination layer. Let $\boldsymbol{x}$ be the input to the network and z_i^+ and z_j^-
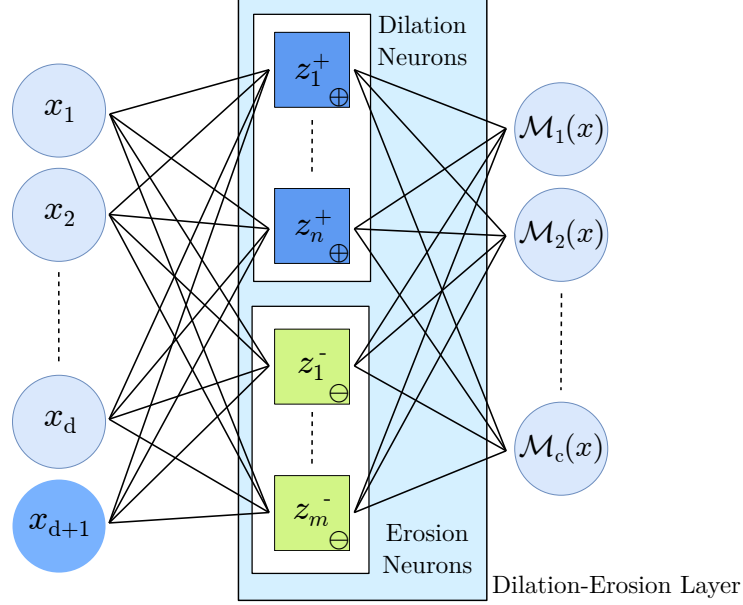
Figure 3.1.: Architecture of single morphological block. It contains an input layer, a dilation-erosion layer with n dilation and m erosion neuron and a linear combination layer with c neurons producing the output.

be the output of the ith dilation neuron and the jth erosion neuron respectively:

$$z_i^+ = \boldsymbol{x} \oplus \boldsymbol{s}_i^+, \tag{3.8}$$

$$z_j^- = \boldsymbol{x} \ominus \boldsymbol{s}_j^-, \tag{3.9}$$

where $\boldsymbol{s}_i^+$ and $\boldsymbol{s}_j^-$ are the structuring elements of the respective neurons. Note that $i \in \{1, 2, \ldots, n\}$ and $j \in \{1, 2, \ldots, m\}$. The final output of a node in the linear combination layer is computed as

$$\mathcal{M}(\boldsymbol{x}) = \sum_{i=1}^{n} z_i^+ \omega_i^+ + \sum_{j=1}^{m} z_j^- \omega_j^-, \tag{3.10}$$

where ω_i^+ and ω_j^- are the weights of the combination layer. When the network is trained, it learns $\boldsymbol{s}_i^+$, $\boldsymbol{s}_j^-$, ω_i^+ and ω_j^-.

3.2.3. Morphological block as sum of hinge functions

Hinging hyperplanes [12] have many applications in regression and classification tasks. In this subsection, we show that the simple morphological block can be represented as a sum of hinge functions. We first define hinging hyper planes and then relate morphological block to hinging hyper planes.

Definition 1 (*k*-order Hinge Function [94])**.** *A k-order hinge function* $h^{(k)}(\boldsymbol{x})$ *consists of*

$(k + 1)$ *hyperplanes continuously joined together. It may be defined as*

$$h^{(k)}(\boldsymbol{x}) = \pm \max\{\boldsymbol{w}_1^T \boldsymbol{x} + b_1, \boldsymbol{w}_2^T \boldsymbol{x} + b_2, \ldots, \boldsymbol{w}_{k+1}^T \boldsymbol{x} + b_{k+1}\} \tag{3.11}$$

Definition 2 (d-order hinging hyperplanes (d-HH) [94])**.** *A d-order hinging hyperplanes (d-HH) is defined as the sum of multi-order hinge functions having the following form.*

$$\sum_i \alpha_i h^{(k_i)}(\boldsymbol{x}), \tag{3.12}$$

with $\alpha_i \in \{-1, 1\}$ and $k_i \leq d$.

The following may be stated about hinging hyperplanes [94].

Theorem 1. *The function computed by a single morphological block (denoted by $\mathcal{M}(\boldsymbol{x})$) with n dilation and m erosion neurons followed by their linear combination, is a sum of multi-order hinge functions.*

Proof. To prove this, it is enough to show that $\mathcal{M}(\boldsymbol{x})$ can be written as the sum of l hinge functions, i.e.,

$$\mathcal{M}(\boldsymbol{x}) = \sum_{i=1}^{l} \alpha_i h_i^{(d)}(\boldsymbol{x}) \tag{3.13}$$

where $l = m + n$, $\alpha_i \in \{1, -1\}$ and $h_i^{(d)}(\boldsymbol{x})$ is i-th d-order hinge function. The detailed proof is given in appendix section. □

3.2.4. Single Morphological block as sum of piecewise linear function

Here we show that $\mathcal{M}(\boldsymbol{x})$ can represent a sum of piece wise linear function. To prove this we have taken the help of following proposition.

Proposition 1 (From [93])**.** *Any continuous piece-wise linear function (PWL) can be expressed as difference between two convex PWL functions.*

Then we show the following.

Lemma 1. $\mathcal{M}(\boldsymbol{x})$ *is a continuous piece-wise linear function.*

Proof. Here is show that $\mathcal{M}(\boldsymbol{x})$ can be written is the following form

$$\mathcal{M}(\boldsymbol{x}) = \sum_{i=1}^{t_1} \phi_i'(\boldsymbol{x}) - \sum_{i=1}^{t_2} \phi_i''(\boldsymbol{x}), \tag{3.14}$$

Where $\phi_i'(\boldsymbol{x})$ and $\phi_i''(\boldsymbol{x})$ are a the PWL function.

The complete proof is given in appendix section. □

3.2.5. Number of hinged hyperplanes

A single morphological block learns the function $\mathcal{M}(\boldsymbol{x})$ as defined in Section 3.2.3. The learned function may also be expressed in the following form.

$$\mathcal{M}(\boldsymbol{x}) = \sum_{i=1}^{l} \alpha_i \max_{k} \{\theta_k x_k + \rho_{ik}\}, \tag{3.15}$$

where α_i, θ_k, $\rho_{ik} \in \mathbb{R}$. The derivation can be found in the supplementary material A.4. We see that $\mathcal{M}(\boldsymbol{x})$ is sum of l functions, each of which computes max over the linearly transformed elements of $\boldsymbol{x}$. Since the max is computed over the (transformed) elements of $\boldsymbol{x}$, each max operation selects only one element of $\boldsymbol{x}$. So, the computed $\mathcal{M}(\boldsymbol{x})$ may not contain all the elements of $\boldsymbol{x}$ and the index (k) of selected element varies depending on the input and the structuring element. However, if $l > d$, $\mathcal{M}(\boldsymbol{x})$ may contain all the elements of $\boldsymbol{x}$. So equation 3.15 can be rewritten as

$$\mathcal{M}(\boldsymbol{x}) = \alpha_1(\theta_1 x_{k_1} + \rho_{1k_1}) + \alpha_2(\theta_2 x_{k_2} + \rho_{2k_2}) + \cdots + \alpha_l(\theta_l x_{k_l} + \rho_{lk_l}). \tag{3.16}$$

where x_{k_i} represents any one of the $d + 1$ elements of $\boldsymbol{x}$ selected by i-th neuron by max operation depending on structuring element $\boldsymbol{s}_i$. So each x_{k_i} is chosen from $d + 1$ elements. Therefore, depending on which element of $\boldsymbol{x}$ gets selected by each neuron, $\mathcal{M}(\boldsymbol{x})$ forms one of the $(d + 1)^l - 1$ hyperplanes. The -1 occurs in number of hyperpalnes because in one occasion all neurons select the augmented 0. Note that some of these hyperplanes must be parallel to some axes. For $\mathcal{M}(\boldsymbol{x})$ to form a hyperplane that is not parallel to any of the axes, all elements of $\boldsymbol{x}$ must get selected by some max functions or other. This occurs in $d! \times \binom{l}{d}$ ways. The remaining $l - d$ number of elements x_{k_i}'s are repeat selection by some functions. So, there can be almost $d! \times \binom{l}{d} \times (d + 1)^{l-d}$ hinging hyperplanes that are not parallel to any of the axes. Therefore, number of hinging hyperplanes increases exponentially with the number of neurons in the dilation-erosion layer. These hyperplanes can act as decision boundaries if the network is employed as a classifier. This is demonstrated experimentally using a toy dataset representing two-class problem.

The toy dataset contains samples that are distributed along two concentric circles, one circle for each class. Suppose, the circles are centered at the origin. We compare the results(Table 3.1) of various networks with two neurons in the hidden layer. It is observed that baseline neural network fails to classify this data as with two hidden neurons it learns only two hyperplanes, one for each neuron. The boundaries learned by the network with ReLU activation function (NN-ReLU) is shown in figure 3.2a. The result of maxout network is better, because, in this case, the network learns $2k = 4$ hyperplanes as shown in figure 3.2b. Note that with two morphological neurons in dilation-erosion layer, our

(a) NN-ReLU (b) Maxout Network (c) Morph-Net (d) Soft Morph-Net
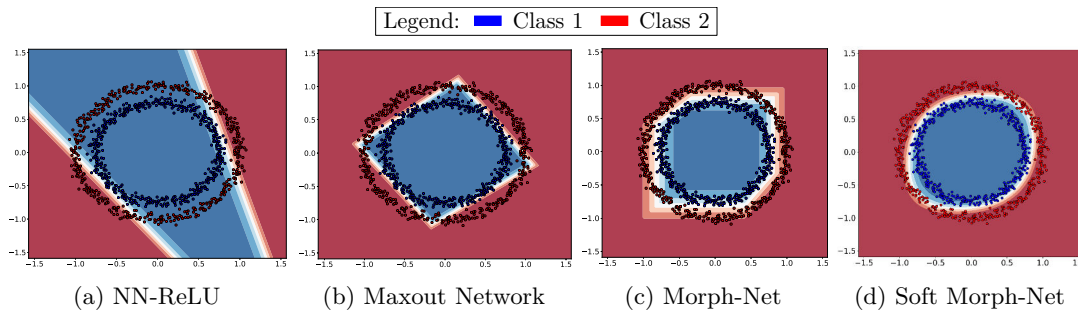
Figure 3.2.: Decision boundaries learned by different networks with two hidden neurons. (a) Baseline neural network is able to learn only two planes (b) Maxout networks is able to learn two more planes with the help of additional parameters. (c) Morph-Net is able to learn more planes with same number of parameters as NN-ReLU. (d) Using soft version of Morph-Net, smooths the learned decision boundary. This further enhances the discrimination capability of the network while retaining the same number of parameters.

Table 3.1.: Training accuracy achieved on the circle dataset by different networks

Methods	Hidden nodes	Parameters	Training accuracy
NN-ReLU	2	12	68.87
NN-tanh	2	12	69.10
Maxout Network (h=2)	2	18	87.17
DenMo-Net	2	12	**91.6**

morphological block has learned 6 hyperplanes to form the decision boundary (figure 3.2c). equation 3.15 suggests that we should get at most 8 distinct lines. However, out of these only two decision boundaries are placed in any arbitrary orientation in the 2D space, while others are parallel to either of the axes.

3.2.6. A single morphological block and universal approximation

A single morphological block represents a sum of hinge functions. However, it is not clear if all hinge functions can be represented by a single morphological block. In a numerical study, we have tried to approximate the hinge function $\max(x + y, 0)$ using a single morphological block by varying the number of dilation/erosion neurons. We have generated values of the function in the square $[-5, 5] \times [-5, 5]$, and trained the network with mean squared error (MSE) loss. In Figure 3.3, we have plotted the MSE loss (after convergence) against the number of morphological neurons used. It is seen that a single morphological block is unable to reduce the error unless we use additional bias or limiter term in the morphological neurons. However, we do not know theoretically if having additional bias terms in morphological operations help in universal approximation.

Unlike using a single morphological block, in the next section we have tried to use two

Loss-without bias and Loss-with bias
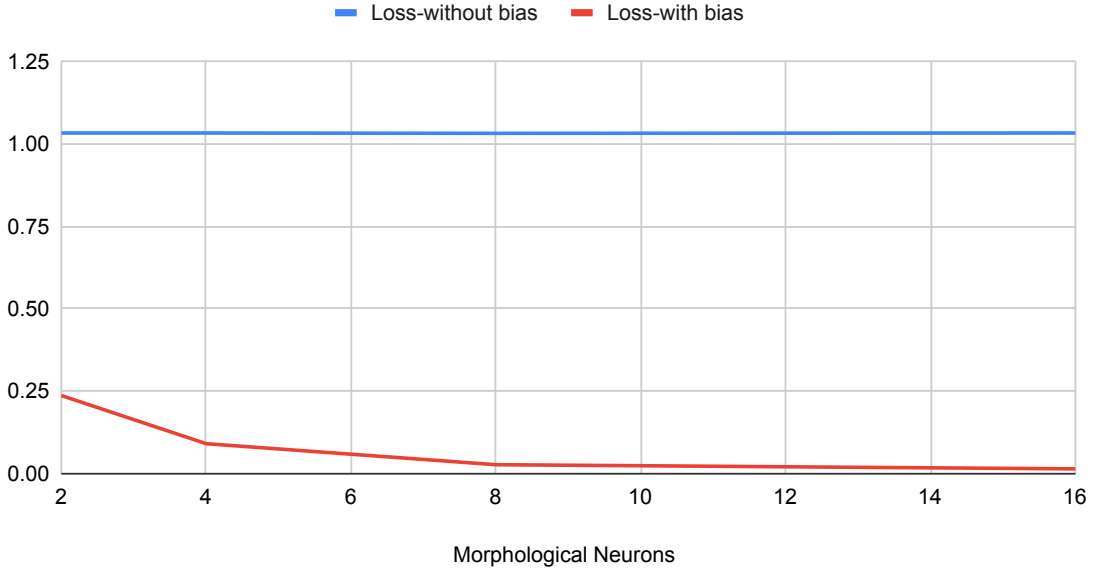
Figure 3.3.: Graph of approximation loss with varying morphological neurons in a single morphological block.

morphological block and proved that it can approximate any continuous function.

3.2.7. Universal Approximation

Here we have tried to show that two-layer Morphological block can approximate any continuous functions. First we have shown that any hyperplane can be represented by a single morphological block. After that we have shown the universal approximation using two morphological block.

Lemma 2. *Let K be a compact subset of $\mathbb{R}^d$. Then, over K, any hyperplane $w^\top \boldsymbol{x} + b$ can be represented as an affine combination of d dilation neurons which only depend on K.*

Proof. Since we are in a compact set, there exists $C > 0$ such that $|x_\ell| \leq C$ for any $1 \leq \ell \leq d$. Where x_ℓ is an each element of $\boldsymbol{x}$. Take

$$s_\ell = -3C\mathbf{1}_d + 3Ce_{\ell,d}, 1 \leq \ell \leq d,$$

where $\mathbf{1}_d$ is the vector of all ones and $e_{\ell,d}$ is the ℓ-th unit vector in $\mathbb{R}^d$. Then all but the ℓ-th coordinate of s_ℓ are $-3C$, while the ℓ-th coordinate is 0. Then note that, for any

45

$\boldsymbol{x} \in K$, and $1 \le \ell \le d$,

$$x_\ell + s_{\ell,\ell} = x_\ell \ge -C > -2C = C - 3C$$
$$\ge x_j - 3C = x_j + s_{\ell,j},$$

for any $j \ne \ell$. It follows that for any $x \in K$, and $1 \le \ell \le d$,

$$\boldsymbol{x} \oplus s_\ell = x_\ell.$$

Now given any hyperplane $w^\top \boldsymbol{x} + b$, we can express it exactly as a linear combination of dilation neurons over K:

$$w^\top \boldsymbol{x} + b = \sum_{\ell=1}^{d} w_\ell x_\ell + b = \sum_{\ell=1}^{d} w_\ell (\boldsymbol{x} \oplus s_\ell) + b.$$

This completes the proof. □

Lemma 3. *Any linear combination of hinge functions $\sum_{i=1}^{m} \alpha_i h^{(k_i)}(\boldsymbol{x})$ can be represented over any compact set K as a two sequential morphological block consisting of dilation neurons only.*

Proof. Let $B = \max_{1 \le i \le m} \sup_{\boldsymbol{x} \in K} |h^{(k_i)}(\boldsymbol{x})|$. We now give the architecture of the desired Morph-Net.

1. The first dilation-erosion layer has exactly d dilation neurons given by $\boldsymbol{x} \oplus s_\ell, 1 \le \ell \le d$.
2. The first linear combination layer has $k = \sum_{i=1}^{m} (k_i + 1)$ neurons, with the i-th block of $(k_i + 1)$ neurons outputting the constituent hyperplanes of $h^{(k_i)}(\boldsymbol{x})$. This can be done by Lemma 2.
3. The second dilation-erosion layer just has m dilation neurons, each outputting a hinge function. The ℓ-th neuron is constructed as follows: Write any $\boldsymbol{y} \in \mathbb{R}^k$ as $(\boldsymbol{y}_1^\top, \ldots, \boldsymbol{y}_m^\top)^\top$ where $\boldsymbol{y}_j = (y_{j,1}, \ldots, y_{j,k_j+1})^\top$. We want the output of the ℓ-th neuron to be $\max_{1 \le v \le k_\ell+1} y_{\ell,v}$. So we take $\boldsymbol{t}_\ell = (\boldsymbol{t}_{\ell,1}^\top, \ldots, \boldsymbol{t}_{\ell,m}^\top)^\top$, where $\boldsymbol{t}_{\ell,j} = -3B \mathbf{1}_{k_j+1}$ for $j \ne \ell$, and $\boldsymbol{t}_{\ell,\ell} = \mathbf{0}_{k_\ell+1}$. Then, for any $j \ne \ell$, $1 \le u \le k_j + 1$, and $1 \le v \le k_\ell + 1$, we have

$$y_{j,u} + t_{\ell,j,u} = y_{j,u} - 3B \le B - 3B$$
$$= -2B$$
$$< -B$$
$$\le y_{\ell,v} = y_{\ell,v} + t_{\ell,\ell,v}.$$

It follows that $\boldsymbol{y} \oplus \boldsymbol{t}_\ell = \max_{1 \le v \le k_\ell+1} y_{\ell,v}$. With this construction, the outputs of the

second dilation-erosion layer are the m numbers $h^{(k_i)}(\boldsymbol{x})$.

4. The second linear combination layer just has a single neuron that combines the outputs of the previous layer in the desired way:

$$\boldsymbol{z} \mapsto \sum_{i=1}^{m} \alpha_i z_i.$$

This completes the proof. □

Theorem 2 (Universal approximation). *Two sequential morphological blocks can approximate continuous functions over arbitrary compact sets.*

Proof. Continuous functions can be approximated over compact sets by sums of hinge functions (see Theorem 3.1 of [12]). Therefore, by Lemma 1, it follows that any continuous function can be approximated over arbitrary compact sets by two sequential morphological blocks. □

3.2.8. Deep Morphological network and its properties

The network we have discussed so far consists of a single dilation-erosion layer followed by a linear combination layer or sequential use of morphological block. Though theoretically two sequential morphological block can represent represent any continuous function, it is worth to know the properties of using such dilation and erosion layers. It has been shown that neural networks with deeper architecture has quite a few advantages [9]. In fact, some functions require deeper architecture for implementation as the deeper layers help in combining features from the previous layers. This potentially allow the data to be modelled using fewer parameters. The same is true for deeper version of Morph-Net, which may be obtained by stacking the layers in one of the following ways.

Type-I Multiple dilation-erosion layers with a single linear combination layer at the end.

Type-II A unit is formed by a Dilation-Erosion layer followed by a linear combination layer, which is then repeated multiple times.

Due to different arrangement of the layers, these networks behave differently. Moreover, for Type-I networks, not all configuration of layers are effective. To analyze theoretically, let us use $D_{m_1}E_{n_1} \rightarrow D_{m_2}E_{n_2} \rightarrow \cdots \rightarrow D_{m_\ell}E_{n_\ell} \rightarrow L$ to denote a Type-I network with ℓ dilation-erosion layers. Here '$D_{m_i}E_{n_i}$' denotes the i-th layer with m_i dilation neurons and n_i erosion neurons, and 'L' denotes the linear combination layer.

Lemma 4. *Suppose f and g are real-valued functions on $\mathbb{R}^d$. Then $f = g$ if and only if, for all $r \in \mathbb{R}$, one has equality of the sub-level sets:*

$$f^{-1}(-\infty, r] = g^{-1}(-\infty, r].$$

Proof. Proof is given in the appendix section. $\square$

We use this lemma to proof the equivalence of architectures for different morphological networks.

Theorem 3 (Equivalence of architectures). *Given two networks of Type-I, the following can be said about their equivalence.*

  (i) *The architecture $D_{m_1}E_0 \to D_{m_2}E_0 \to \cdots \to D_{m_\ell}E_0$ consisting only of dilation layers is equivalent to the architecture $D_{m_l}E_0$ with a single dilation layer. A similar statement is true if one considers architectures with purely erosion layers.*

 (ii) *The architecture $D_1E_1 \to D_1$ is not equivalent to D_1E_0. Similarly, it is not equivalent to D_0E_1, and consequently, the architectures $D_1E_1 \to D_1E_1$ and D_1E_1 are not equivalent.*

(iii) *The architecture $D_1E_1 \to D_1 \to L$ is not equivalent to $D_1E_0 \to L$.*

(iv) *The architecture $D_2E_0 \to D_0E_2 \to D_1$ is not equivalent to $D_2E_0 \to D_1$.*

Proof. Proof is given in the appendix section. $\square$

For Type-II networks, the similar analysis is difficult due to existence of linear combination layer in between the dilation-erosion layers. However, in the previous section we have shown the approximation properties for Type-II networks.

As stated earlier, use of deep Morph-net may prove beneficial. However, the use of multiple layers amplifies the problem of gradient propagation further. But, as we will see in the next section, this problem is almost alleviated in 2D Morph-Net as the size of structuring element is usually much less than the input or the output. The same problem of gradient propagation occurs if the structuring element is large. This may happen in classifier networks where the size of the feature map is reduced in each stage before the final classification.

3.3. Experimental results

In this section we present some applications of our proposed Morph-Net and empirically evaluate the performance of the same and its variants. It is already shown that a single

Table 3.2.: Accuracy on MNIST and Fashion-MNIST Datasets using a single hidden layer with 400 morphological neurons.

Dataset	Test Accuracy		
	Morph-Net	Soft Morph-Net $(\beta = 8)$	State-of-the-art
MNIST	98.39	98.90	**99.79** [92]
Fashion-MNIST	89.87	**89.84**	89.70 [98]

layer Morph-Net represents a sum of hinge hyper planes. So we may apply this proposed network to realize decision boundaries of classifier. To evaluate the performance of our Morph-Net, we apply it on various benchmark datasets like MNIST [40], Fashion-MNIST [98], SVHN [57] and CIFAR-10 [39]. We have compared the results with that of similar structured (i.e., fully connected) neural network using various activation functions, such as tanh (**NN-tanh**) and ReLU (**NN-ReLU**). We have also chosen the maxout network for comparison, particularly because it uses the *max* function as a replacement of the activation function.

In all the experiments, whenever necessary, the data is flattened in row major order and normalized to the range [0, 1], before feeding to the network. So, the methods can not exploit the spatial structures present in the image data. In all the networks, *softmax* is used as activation function in the last layer. Trainable parameters are initialized using the method of Glorot and Bengio [27], except for the bias which is initialized to zero. In the Morph-Nets, same number of dilation and erosion neurons are used unless stated otherwise. All the networks have been trained using categorical cross entropy loss and Adam optimizer (learning rate=0.001, β_1=0.9, β_2=0.999) [37]. Note that the value of β for computing soft maximum is chosen empirically. The results are presented in the following subsections.

3.3.1. MNIST Dataset

MNIST dataset [40] contains gray scale images of hand written numbers (0-9) of size 28×28. It has 60,000 training images and 10,000 test images. Since our network does not support two dimensional input, we have converted each image into a column vector (in row major order) before giving it as input. The network we use follows the structure we have defined previously: input layer, dilation-erosion layer and linear combination layer computing the output. As in this dataset we had to distinguish between 10 classes of images, 10 neurons are taken in the output layer.

In table 3.2 we have shown the accuracy on test data after training the network for 150 epochs with different number of nodes (l) in the dilation-erosion layer. The change of test accuracy over the epochs is shown in figure 3.4a. It is seen that increasing number of nodes

(a) Varying l in the dilation-erosion layer

(b) Using only dilation, only erosion and both
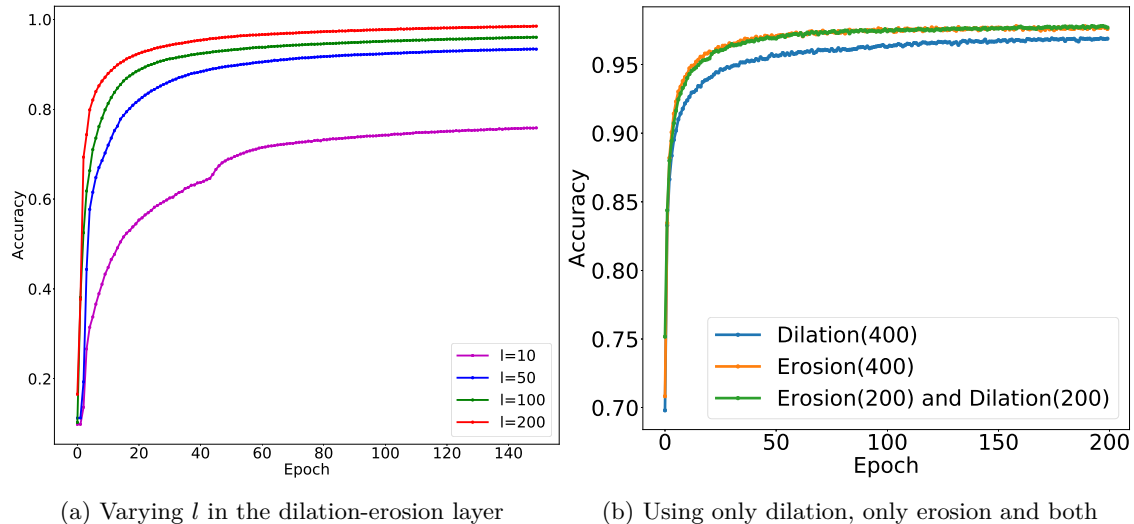
Figure 3.4.: Test accuracy achieved over epochs in the MNIST dataset

in the dilation-erosion layer helps to increase non-linearity, and thus it results in better accuracy on test data. We get test average accuracy of 98.43% after training 3 times with the DenMo-Net of 200 dilation and 200 erosion neurons (Table 3.2) up to 400 epochs. We have also experimented when only dilation neuron, only erosion neuron and both type of neurons were used in the dilation-erosion layer (Figure 3.4b). We see that using both erosion only and both dilation and erosion neurons giving better accuracy. Note that the performance of our proposed Morph-Net is comparable with state-of-the-art performance.

3.3.2. Fashion-MNIST Dataset

The Fashion-MNIST dataset [98] has been proposed with the aim of replacing the popular MNIST dataset. Similar to the MNIST dataset this also contains 28×28 images of 10 classes and 60,000 training and 10,000 testing samples. While MNIST is still a popular choice for benchmarking classifiers, the authors' claim that MNIST is too easy and does not represent the modern computer vision tasks. This dataset aims to provide the accessibility of the MNIST dataset while posing a more challenging classification task.

For the experiment, we have converted the images to a column vector similar to what we have done for the MNIST dataset. We have taken 400 dilation and 400 erosion nodes in the dilation-erosion layer for this experiment. We have trained the network separately 3 times up to 300 epochs. The reported test accuracy (Table 3.2) is the average of the 3 runs. We see that our method gives results comparable to that of state-of-the-art method.

(a) CIFAR-10   (b) SVHN
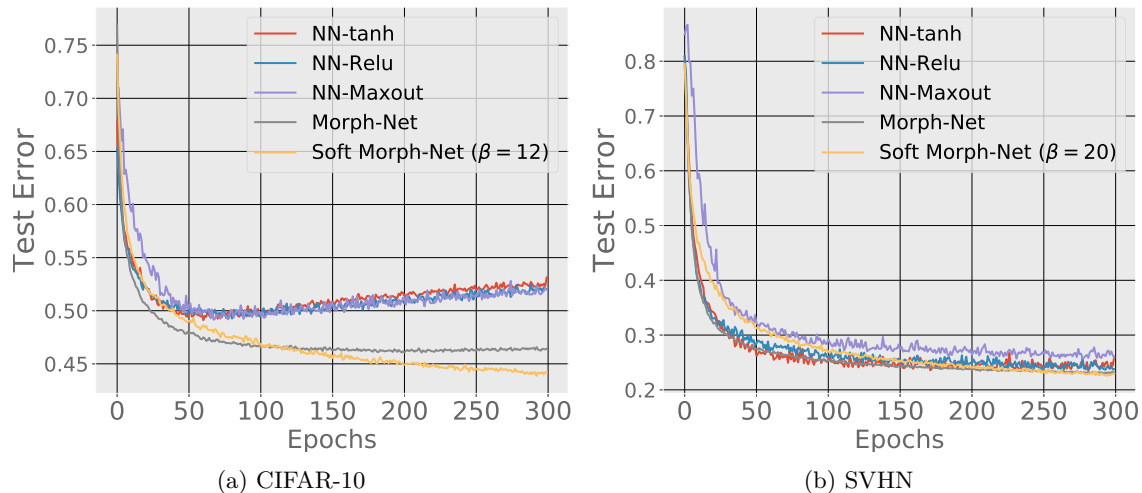
Figure 3.5.: Test accuracy attained over the epochs by different methods.
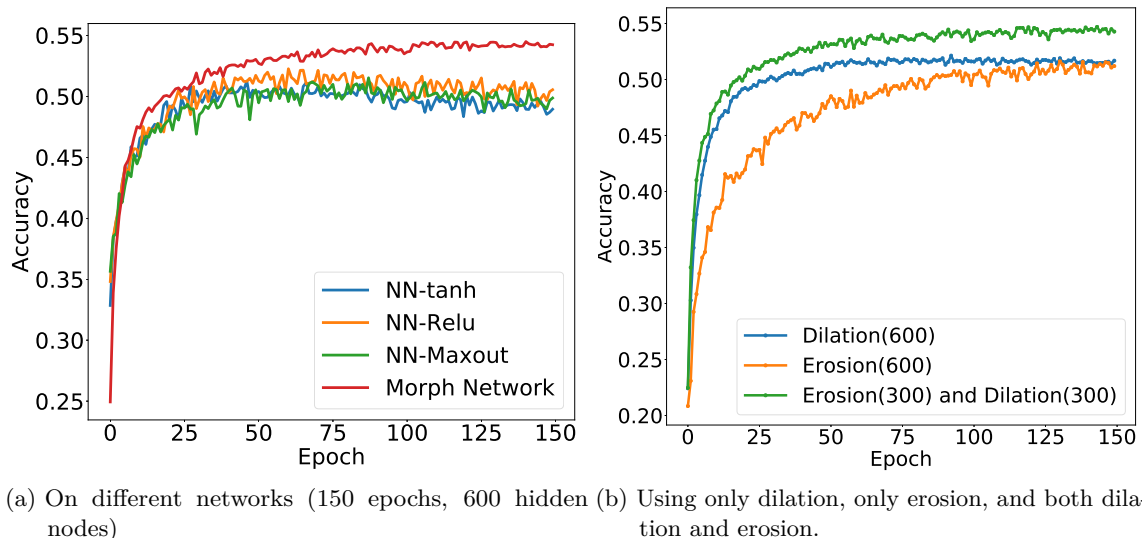
(a) On different networks (150 epochs, 600 hidden nodes)

(b) Using only dilation, only erosion, and both dilation and erosion.

Figure 3.6.: Test accuracy over epochs on CIFAR-10 dataset by morphological block

Table 3.3.: Test accuracy achieved on CIFAR-10 and SVHN dataset by different networks when the number of neurons (l) in the hidden layer is varied. The value of β is taken 12 and 20 for CIFAR10 and SVHN respectively.

Architecture	l=200		l=400		l=600	
	CIFAR10	SVHN	CIFAR10	SVHN	CIFAR10	SVHN
NN-tanh	46.6 ± 0.06	73.9 ± 0.12	46.9 ± 0.04	73.9 ± 0.23	48.0 ± 0.05	75.6 ± 0.14
NN-ReLU	47.2 ± 0.11	64.2 ± 0.88	48.0 ± 0.05	76.2 ± 0.32	48.1 ± 0.02	$\mathbf{79.5 \pm 0.11}$
Maxout-Network ($k = 2$)	46.9 ± 0.05	69.4 ± 0.10	48.0 ± 0.10	74.1 ± 0.22	46.4 ± 0.33	37.8 ± 3.15
Morph-Net	52.0 ± 0.02	73.4 ± 0.03	53.6 ± 0.01	76.9 ± 0.03	54.0 ± 0.02	78.2 ± 0.03
Soft Morph-Net ($\beta = 12, 20$)	$\mathbf{53.5 \pm 0.04}$	$\mathbf{74.1 \pm 0.06}$	$\mathbf{55.8 \pm 0.05}$	$\mathbf{77.0 \pm 0.05}$	$\mathbf{56.9 \pm 0.04}$	78.5 ± 0.05

(a) CIFAR-10　　　　　　　　　　　(b) SVHN
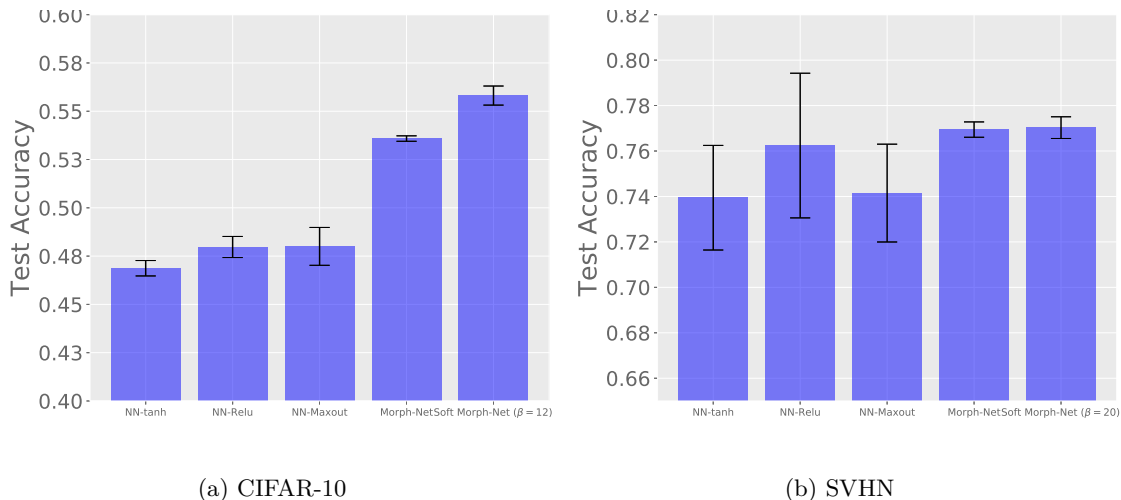
Figure 3.7.: Mean and standard deviation of accuracy achieved by different methods on the test set over 5 runs.

3.3.3. CIFAR-10 and SVHN

CIFAR-10 [39] is a natural image dataset with 10 classes. It has 50,000 training and 10,000 test images. SVHN dataset contains images of numerals written in English, collected from house numbers in Google Street View images. The dataset has 73257 training samples and 26032 test samples. Both datasets contain color images of size 32×32.

For evaluating the performance of the networks considered here, all of them are made to have same number of neurons in the hidden layer. For maxout network, we have taken $k = 2$ that means output of each hidden neurons is calculated without any activation function, but by taking maximum over output from two neurons. Table 3.3 shows the mean and standard deviation of test accuracy obtained over 5 runs of 300 epochs each and by varying the number of neurons in the hidden layer. The change of average test errors over the epochs are shown in figure 3.5a and figure 3.5b for the respective datasets when number of hidden neurons is 400. We have also shown the change of accuracy with 600 hidden neurons is 600 in figure 3.6a. For CIFAR-10 dataset, using only a single type of neurons in our network, we see a different result(Figure 3.6b). The network takes time to learn with only erosion neurons. The situation improves a little when only dilation neurons are used. When using both dilation and erosion neurons, the proposed network perform best by leveraging the power of both the operations. It is seen from Table 3.3 and the figures that our Morph-Net achieves better accuracy for CIFAR-10 dataset in all cases. However, for SVHN dataset, its results are comparable with that of other networks. Note that we have used $\beta = 12$ for CIFAR-10 and $\beta = 20$ for SVHN. Lower value of β results in lower accuracy.

Table 3.4.: Test accuracy achieved on Higgs dataset

Architecture (l=200)	Morph-Net	Soft Morph-Net ($\beta = 20$)	NN-tanh	NN-Relu	NN-Maxout
Accuracy	73.23	72.84	71.20	74.34	**74.88**

3.3.4. Higgs Dataset

Higgs Dataset [7] is built to benchmark the performance of neural networks in distinguishing signal process producing Higgs boson from the background process that does not. This is a synthetically generated dataset with 28 features commonly used by the physicists to distinguishing between the two. The dataset has 11 million data instances. Out of which we have taken random 80% as the training data and rest as the test data. The features have been normalized between -1 and 1 before they are sent to training. In Table 3.4, we have reported the performance of the network. We see that Morph-Net performs better than Soft Morph-Net. However, the performance of the proposed network and that of other networks are very similar.

3.4. Summary

In this Chapter, we have proposed a novel trainable morphological network (Morph-Net) using dilation and erosion operators as neurons. It is shown that these operators, in conjunction with linear combinations, represent a sum of hinge functions. As for multi-layer morphological networks, we have proved that taking two sequential morphological blocks can approximate any continuous function. Note that max (resp. min) function associated with dilation (resp. erosion) operation, being only piece-wise differentiable, allows updating only a single component of the structuring element. This slows down the training of the network. To overcome this problem, we have introduced soft-max (and soft-min) in dilation (resp. erosion) neurons. We have also explored the properties of morphological network for multiple layers. We have explored the theoretical aspects of the proposed morphological network in 1D. It can be readily extended to 2D morphological network forming a CNN like network. In the next chapter we have explored 2D morphological network with all the characteristics of 1D. Efficacy of 2D Morph-Net is evaluated by applying on various computer vision problems.

Part IV.

2D Morphological Network

Chapter 4.

2D Morphological Network

4.1. Introduction

In the previous chapter we have analyzed the network to process data with 1D structuring elements. In this Chapter we have extended the idea of 1D morphological network to 2D morphological network. It is noted that, in the recent years, deep convolutional neural network (DCNN) has significantly outperformed the traditional methods of image analysis. This also be noted that the most basic morphological operation dilation and its dual, erosion, have quite a bit of similarity with convolution operation. Similar to convolution, these morphological operators work by operating a structuring element (kernel in case of convolution) locally over the whole input. However, instead of sum of products, dilation (or erosion) compute max (resp. min) over the sum (resp. difference). More complex morphological operators are constructed by the composition of these two basic operators. If we draw an analogy with convolutional neural networks (CNN), there also a sequence of convolution operators are employed to extract hierarchy of features. The network starts from simple ones, like edges and corners, and progressively composes them to extract complex features. Learning the hierarchy of features automatically from data is one of the reasons of the success of DCNNs. Similar things may be achieved if we arrange the dilation and erosion operations in a network form and try to learn the structuring elements automatically from data. In the previous chapter we have shown that a dilation-erosion layer followed by a linear combination layer represents a sum of hinge hyper planes. We extend this novel idea to the 2D cases using 2D morphological blocks, where the dilation and erosion are performed locally. We have employed 2D morphological blocks repeatedly in the network to perform different tasks.

2D morphological networks are made to process gray or color images in efficient way. We have evaluated the performance of the network on various image processing tasks like classification of aerial images, segmentation of blood vessel from fundus images, segmentation of lungs from chest x-ray, document image binarization and image dehazing. The results show that using simple architectures, 2D Morphological network is able to

achieve comparable performance with CNN if not better. Note that number of parameters we need to train is significantly less than the competitive methods. We have also tried to do image classification using 2D Morphological Network. We have chosen aerial images for the classification task using Morphological Network.

There are many different approaches for retinal vessel segmentation.One of the major problem to segment retinal blood vessels is that these are too small to identify. [86] used two-dimensional Gabor wavelet transform responses taken at multiple scales to classify each image pixel as vessel or non-vessel. Marín et.al. [46] performs supervised classification with the help of neural network. In a separate work [71] the basic line detector for pixel classification. In [87] feature vectors are computed from patches and the line elements from retinal color images and classified using a kNN-classifier. On the other hand lungs segmentation is another segmentation problem where we try to segment large blob in an image. Candemir et.al. [14] tried to detect lung boundaries by nonrigid registration-driven method for lung segmentation There are few Deeplearning based methods for lungs segmentation. Rashid et.al. [67] applied fully convolutional neural network to segment out lungs from chest x-ray images. In a separate work [36] the Encoder-Decoder architecture is used.

Document image binarization and elimination of noise is an important task in document iamge processing. However, this is more challenging in case of historical documents, because of varying degradation types. There are various classical binarization methods available which compute local and global thresholds for binarization [61], [78]. Dubois and Pathak [18] have proposed a model which is capable of reducing the *bleed through noise* found in graylevel document images due to back impression while scanning. Moghaddam and Cheriet [51] have proposed a restoration method for single-sided low-quality document images (RSLDI). Rivest-Hénault *et. al.* [73] have proposed a local linear level set method for binarization and restoration of degraded document images. Jia *et. al.* [35] have proposed a model of restoration and binarization of historical documents by analyzing the structural symmetry of the strokes. Although there are multiple symbols involved in document image, Other than classification and segmentation we have also shown that morphological network can also perform image dehazing. The objective is to determine accurately determine the amount of haze the images as well as restore the color. Related work regarding image de-hazing is discussed in chapter 2.

Rest of the chapter is organized as follows. In section 4.2, we have described the most basic form of our 2D morphological network and also how it is arrived from 1D morphological network. In the next section, i.e., in section 4.3 we have shown how this network is applied to solve various problems with different architectures of Morphological network. Finally, concluding remarks are presented in Section 4.4.

$$\min\{0{-}0,\ 0{-}1,\ 0{-}1,\ 0{-}2,\ 1{-}1,$$
$$0{-}2,\ 0{-}(\text{-}3),\ 3{-}1,\ 1\text{-}3\} = \text{-}2$$
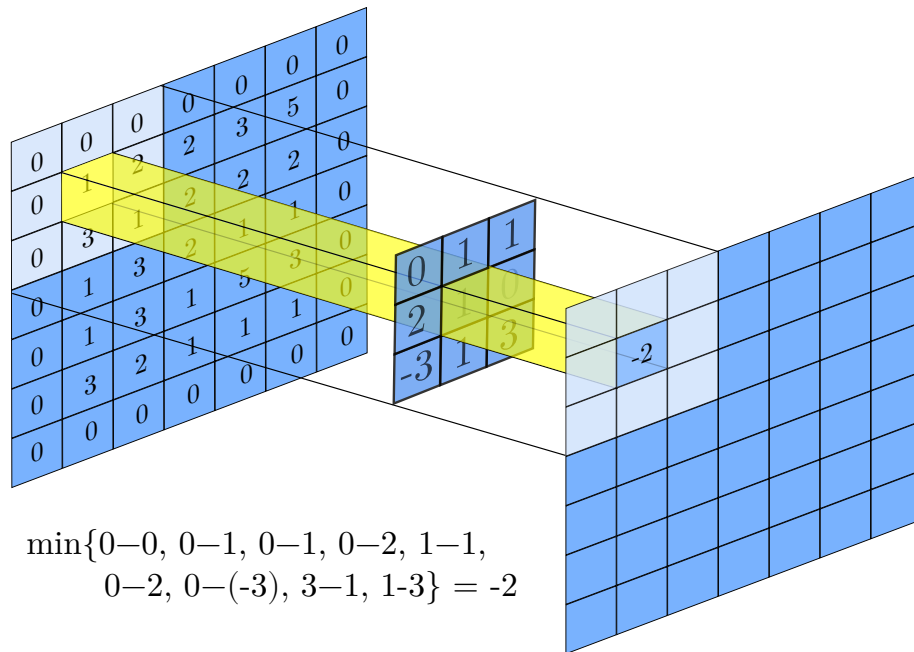
Figure 4.1.: Erosion operation performed with a structuring element of size 3×3. This generates a erosion feature map.

4.2. 2D Morphological Network

In this section we recall the basic dilation and erosion operation as we have already defined in the second chapter. Next we define a 2D morphological block using dilation and erosion operation. Subsequently, we build a network using this morphological block and apply on different problems.

4.2.1. 2D Dilation and Erosion

In the previous chapter we have defined that the dilation and erosion neurons work with structuring elements of same size as the input. However, we may also consider the said input vector is obtained by flattening a part of an image or, more specifically, a small block of the image around a particular pixel. This suggests that all the properties of 1D morphological network holds for 2D network as well. So, in general, these morphological operations can be applied spatially on local neighborhood (called *receptive field*) around each pixel by utilizing structuring elements of appropriate size. In this case, the dilation ($\oplus$) and erosion ($\ominus$) operations on an image I may be defined as follows.

$$(I \oplus S^+)(x,y) = \max_{(i,j) \in D} \{I(x-i, y-j) + S^+(i,j)\} \tag{4.1}$$

$$(I \ominus S^-)(x,y) = \min_{(i,j) \in D} \{I(x+i, y+j) - S^-(i,j)\} \tag{4.2}$$

where $S^+, S^- \in \mathbb{R}^{a \times b}$ are the structuring elements of dilation and erosion operator respectively. $D = D_1 \times D_2$ is the domain of structuring elements, where $D_1 = \{1, 2, \ldots, a\}$ and $D_2 = \{1, 2, \ldots, b\}$. These operations are applied on the whole image by spatially sliding the structuring element, similar to what is done in case of convolution operation (Figure 4.1). Use of the sliding neighborhood operation, in general, helps in extracting spatial features from the image. This is true here also.

Let us call the output generated by dilation operation as **dilation feature map** and that by erosion operation as **erosion feature map**. These feature maps are linearly combined to form a single layer **2D Morph-Net**. It may be noted that, each position (obtained by sliding) of structuring element produces an output and corresponding gradient values affect the same structuring element.

4.2.2. 2D Morphological Block

Let I denotes input feature map of size $(m \times n \times c)$. Let $W_d \in R^{a \times b \times c}$ and $W_e \in R^{a \times b \times c}$ are dilation and erosion kernels or SEs respectively of size $a \times b \times c$ that are applied I. It may be noted that $a \leq m$ and $b \leq n$. Dilation ($\oplus$) and erosion ($\ominus$) operations on image I are defined as the following

$$(I \oplus W_d)(x,y) = \max_{i \in S_1 j \in S_2 k \in S_3} (I(x+i, y+j, k) + W_d(i,j,k)) \tag{4.3}$$

$$(I \ominus W_e)(x,y) = \min_{i \in S_1 j \in S_2 k \in S_3} (I(x+i, y+j, k) - W_e(i,j,k)) \tag{4.4}$$

where $S_1 = \{1, 2, .., a\}$, $S_2 = \{1, 2, .., b\}$ and $S_3 = \{1, 2, .., c\}$. In Figure 4.1 we have graphically shown the working of erosion operation. If c_1 number of dilation and c_2 erosion kernels are used then the output feature map I_1 has size $(m \times n \times c_3)$, where $c_3 = c_1 + c_2$. Note that in our task we have always used zero padding.

In the next layer, the feature maps of I_1 are combined linearly as

$$I_2(x,y) = b + \Sigma_{k=1}^{c_3} w_k I_1(x,y,k) \tag{4.5}$$

where w_k are the weights of the linear combination and b is bias. If λ be the number of such linear combinations, output size would be $(m \times n \times \lambda)$. We call dilation-erosion layer
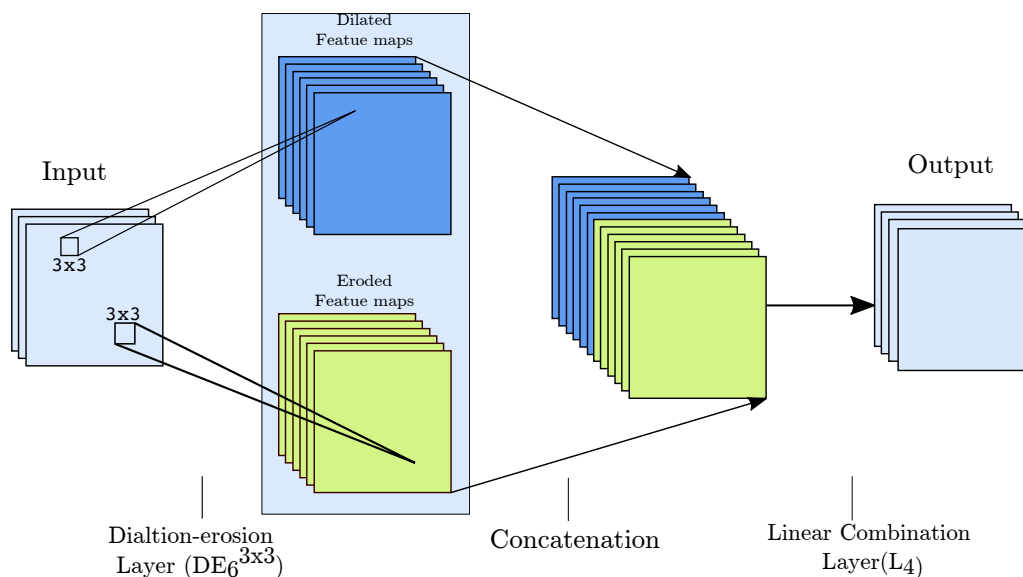
Figure 4.2.: A single 2D morphological network block consisting of a Dilation-Erosion layer and concatenation of its output followed by multiple linear combinations.

followed by linear combination layer together a *2D Morphological block*. Figure 4.2 shows such a block with $a = b = 3$, $c_1 = c_2 = 6$ and $\lambda = 4$. It consist of a dilation-erosion layer with 6 dilation and 6 erosion feature map, followed by linear combination layer with 4 feature map. This entire block is denoted by $DE_6^{3 \times 3} - L_4$. We have shown in previous chapter that a dilation-erosion layer followed by a liner combination layer represents a sum of hinge hyperplanes. The motivation to build 2D morphological block by taking this two layers (dilation-erosion and linear combination) consecutively, because it exactly matches basic principle of dense morphological network. On the other hand, applying dilation and erosion on image reduces dark and bright noise respectively. Their linear combination results in removal of both kinds of noise. In the next section we show how series of 2D morphological blocks can be useful to build a network.

4.2.3. 2D Morphological Network Construction

In the previous section we have defined 2D morphological block. Now, we can build a 2D morphological network for image processing task with the help of 2D morphological block. Let us assume input image size is $M \times N$. Since $a << M$ and $b << N$ at each location (i, j) of image X we consider a box of pixels of size $a \times b$ which is operated on by an structuring element to produce the output at (i, j) independent of output at any other location. The receptive field size is $a \times b$ for taking only a single morphological block. To increase the receptive field we have to take multiple morphological block. It helps to make the decisions by compounding blocks' output from the whole image. In the following

section we have applied the proposed 2D morphological network to different applications in image processing.

4.3. Applications

In this section we present some applications of our proposed Morph-Net and empirically evaluate the performance of the same and its variants. The network is evaluated on the tasks that require image to be taken as input. More specifically, we address the problems of (i) classification of Aerial image, (ii) segmentation of blood vessel in fundus image, (iii) segmentation of lung from X-ray images (iv) binarization of old document imagesg and (v) image dehazin. Although the results are reported on a handful problems only, Morph-Net can be applied to a variety of other image processing problems.

4.3.1. Aerial image Classification

Aerial image classification is a challenging problem. Each image has adequate structural features. Morphological structuring elements are well known for finding structures in the images. We have done aerial image Classification by using proposed 2D morphological network. To do that we have chosen Alex-net [40] type architecture, where each convolution layer is replaced with a 2D morphological block.

(a) Agricultural   (b) Airplane   (c) Buildings   (d) Denseresidential   (e) Intersection
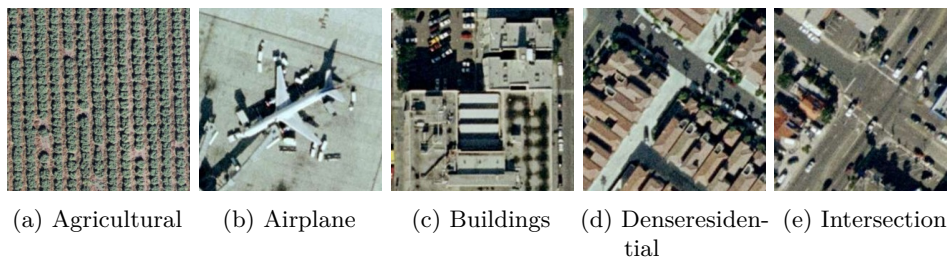
Figure 4.3.: Samples from UC Merced dataset.

Experiments

We have applied 2D morphological network for image classification task on two publicly available datasets named UCMerced and WHU-RS19. UCMerced Land-use Dataset [99] has 2,100 aerial scene images of size 256×256. Few samples of the dataset are shown in Figure 4.3. There are 21 classes in the dataset namely agricultural, airplane, base- ball diamond, beach, buildings, chaparral, dense residential, forest, freeway, golf course, harbor, intersection, medium density residential, mobile home park, overpass, parking lot, river, runway, sparse residential, storage tanks, and tennis courts. It may be noted that the

(a) Airport　　(b) Bridge　　(c) Commercial　(d) FootballField　(e) Parking
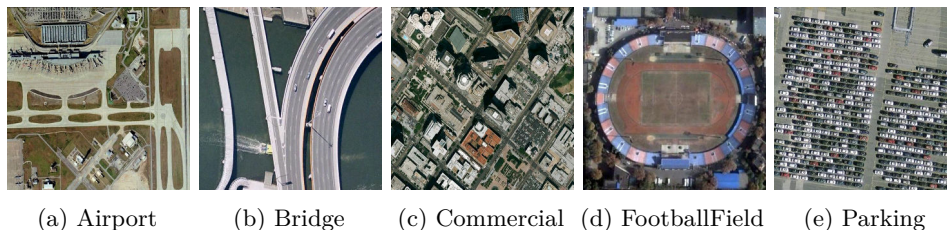
Figure 4.4.: Samples from WHU-RS19 dataset.

Table 4.1.: Accuracy achieved on UC Merced and WHU-RS19 Dataset

Method	Parameters (Millions)	Accuracy	
		UCMerced	WHU-RS19
AlexNet-based	4.42	53.29 ± 0.86	48.26 ± 2.83
Depth-LeNet	5.04	54.81 ± 1.25	47.19 ± 2.43
DeepMorphLeNet	6.04	56.52 ± 1.74	52.91 ± 2.60
AlexNet-based	6.5	72.62 ± 1.05	64.38 ± 2.93
Depth-AlexNet-based	7.47	73.14 ± 1.43	63.27 ± 2.14
DeepMorphAlexNet[59]	10.5	$\mathbf{76.86 \pm 1.97}$	68.20 ± 2.75
Morph-Net(ours)	**2.23**	74.75 ± 1.16	$\mathbf{71.23 \pm 2.91}$

dataset has highly overlapping classes such as the dense, medium, and sparse residential classes which mainly differ in the density of structures.

WHU-RS19 Dataset [96] has 1,005 high-resolution images of 600×600. The dataset contains 19 classes that are airport, beach, bridge, river, forest, meadow, pond, parking, port, viaduct, residential area, industrial area, commercial area, desert, farmland, football field, mountain, park and railway station. The dataset is collected from different regions all around the world. It creates challenges due to variation in scale, orientation, and the illumination of the images. Few samples of the images are shown in Figure 4.4.

For training and testing with 2D morphological network, we have utilized the configuration mentioned in [59]. We have employed Alex-net like architecture for this dataset. The network trained in each fold for upto 300 epochs for each dataset. All the images are resized to 256×256 before it is feed into the network. From the results reported in the Table 4.1, it can be seen that our network works as good as the method of [59].
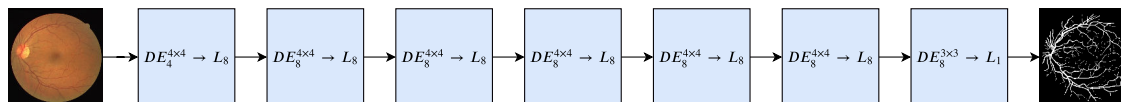
Figure 4.5.: Architecture of the Morph-Net utilized for blood vessel segmentation in retinal images.

(a) input   (b) Morph-Net   (c) Ground Truth
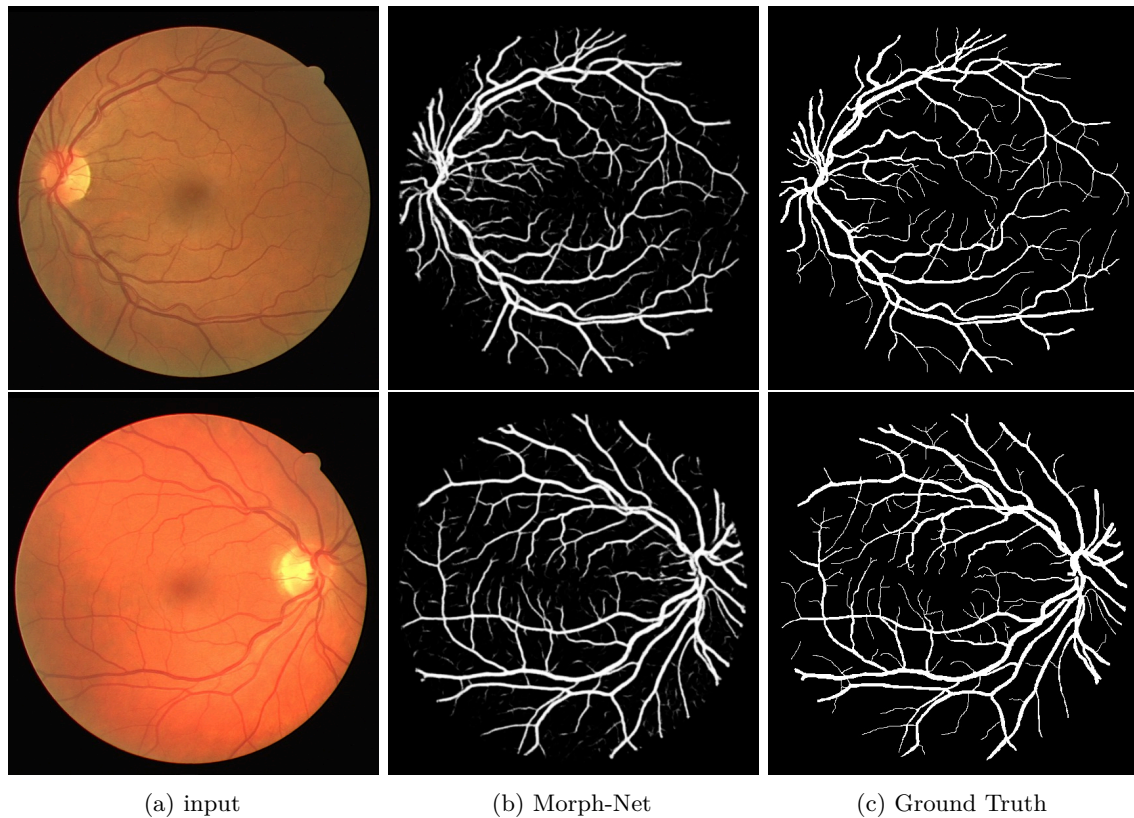
Figure 4.6.: Results obtained by 2D Morph-Net on two samples of DRIVE dataset

Table 4.2.: Accuracy and AUC achieved on the test set by different networks on Drive and STARE dataset.

Method	DRIVE		STARE	
	Accuracy	AUC	Accuracy	AUC
Marin et al. [46]	94.50	0.958	95.20	**0.976**
Staal et al. [87]	94.40	0.952	-	-
Niemeijer et al. [58]	94.10	0.929	-	-
Soares et al. [86]	94.60	0.961	94.80	0.967
Ricci et al. [71]	95.90	0.963	**96.40**	0.968
2D Morph-Net	**96.50**	**0.977**	93.40	0.937

4.3.2. Blood Vessel Segmentation in Retinal Images

Retinal blood vessel segmentation is an important task as it plays an important role in automatic retinal disease screening systems. It is noted that these blood vessels are very sharp and doesn't follow any regular geometrical shapes. To segment these Blood vessels we have used proposed Morphological network as shown in Figure 4.5. We have used 7 sequential block of 2D morpholoigcal block. Each block comprises 8 dilation and 8 erosion feature maps followed by 8 linear combination modules. In the Last layer we have generated output by taking a single linear combination. It maybe noted that we do not employ any other activation function except the last layer to produce output between 0 and 1.

(a) input (b) Morph-Net (c) Ground Truth
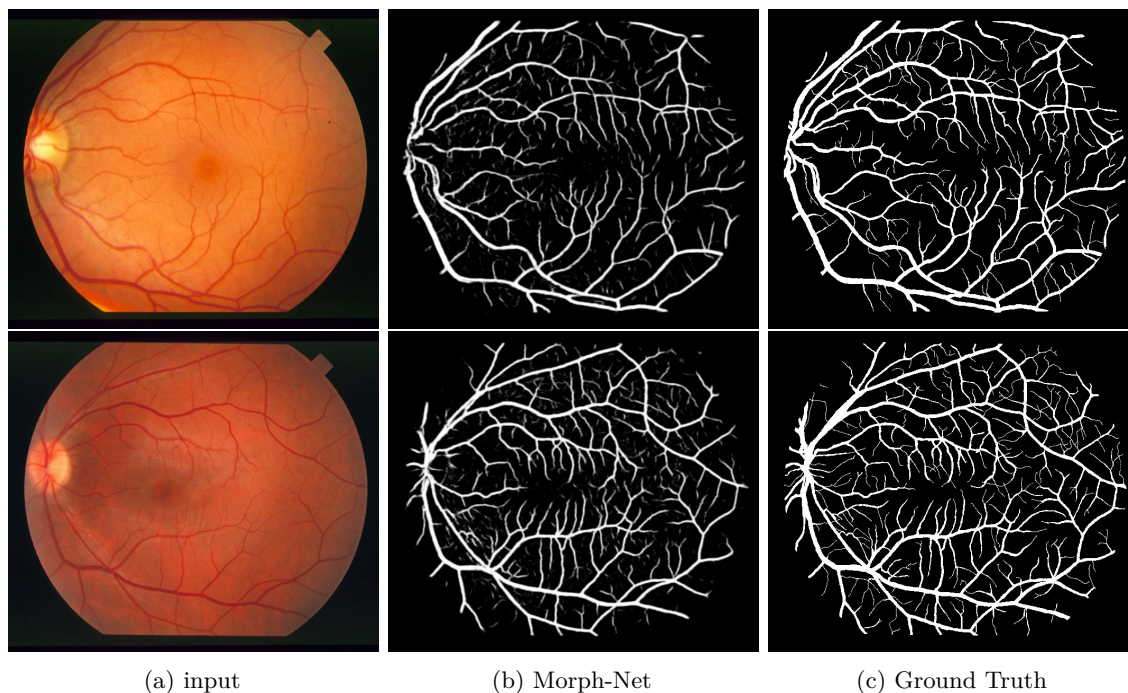
Figure 4.7.: Results obtained by 2D Morph-Net on two samples of STARE dataset.

Experiments

To study the performance on this task, we have used two publicly available retinal image dataset: DRIVE[1] and STARE[2] [33]. In each image of the datasets, vascular structure has been precisely marked by experts. The DRIVE dataset has 40 eye-fundus color images captured with Canon CR5 non-mydriatic 3CCD camera with a 45° field-of-view (FOV). Each image is of size 768×584. The dataset is divided into two groups: 20 training and 20 test images. The STARE dataset consists of 20 eye-fundus color images captured with

---

[1]https://drive.grand-challenge.org/
[2]http://cecas.clemson.edu/ ahoover/stare/

a TopCon TRV-50 fundus camera at 35° FOV. Each of them has a size of 700×605. The network is designed to take an input of size $512 \times 512 \times 3$ and produces a output of size 512×512. So, database images are resized to 512×512 using bilinear interpolation, before they are fed to the network. The network is trained using the binary cross-entropy loss and Adam optimizer [37] with batch size 6. The images obtained from the network is thresholded at 0.5 to get the final result. Since, unlike DRIVE dataset, STARE dataset is not divided in training and testing sets. So we train the network using the training set of the DRIVE dataset till convergence, and test the network using test set of the DRIVE data as well as all the images of STARE dataset. Table 4.2 presents the results in terms of accuracy and area under the curve (AUC). We measure the accuracy as

$$Accuracy = \frac{TP + TN}{TP + FN + TN + FP} \tag{4.6}$$

where TP, TN, FP, FN are true positive, true negative, false positive, false negative respectively. The Table 4.2 shows that the network performs well also on the STARE dataset even being trained on DRIVE dataset. For both the datasets Morph-Net performs comparably, if not better. For qualitative comparison, we have presented two samples from each dataset in figures4.6 and 4.7. Note that extracted blood vessels are little thicker in a few places in the images of DRIVE (figure 4.6). This occurs due to use of fixed threshold 0.5 in the final step.
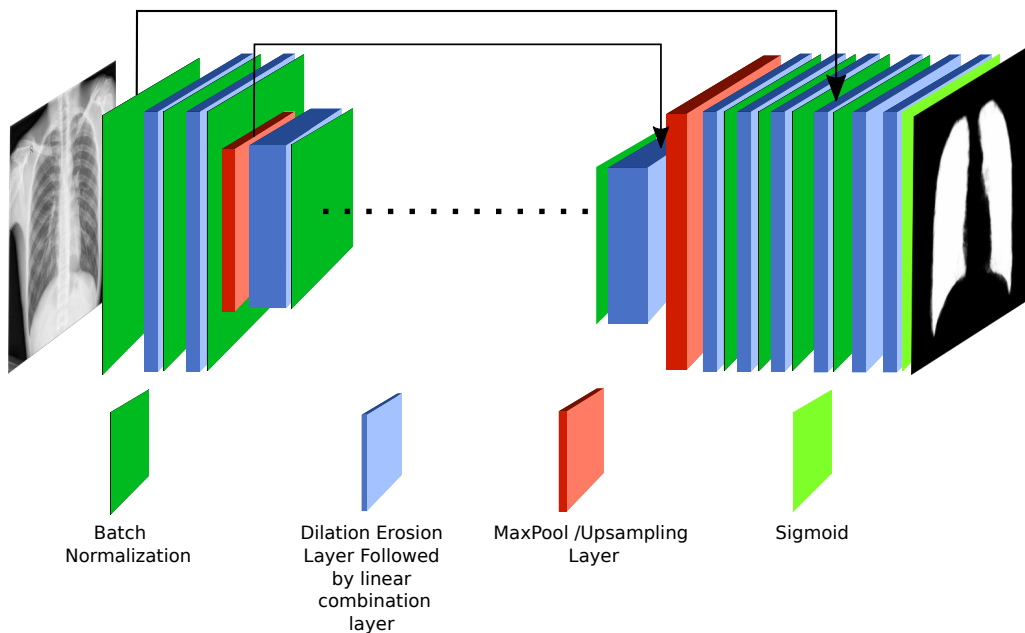
Figure 4.8.: U-Net like architecture of the 2D Morph-Net utilized for lungs segmentation.

Table 4.3.: Attained Dice coefficient on Shenzhen dataset by different networks.

Method	Candemir et al.[14]	ED-CNN[36]	FCN[67]	2D Morph-Net
Dice Coefficient	94.1	97.4	97.7	95.9

(a) input   (b) Morph-Net(before thresolding)   (c) Morph-Net(after thresolding)   (d) Ground Truth
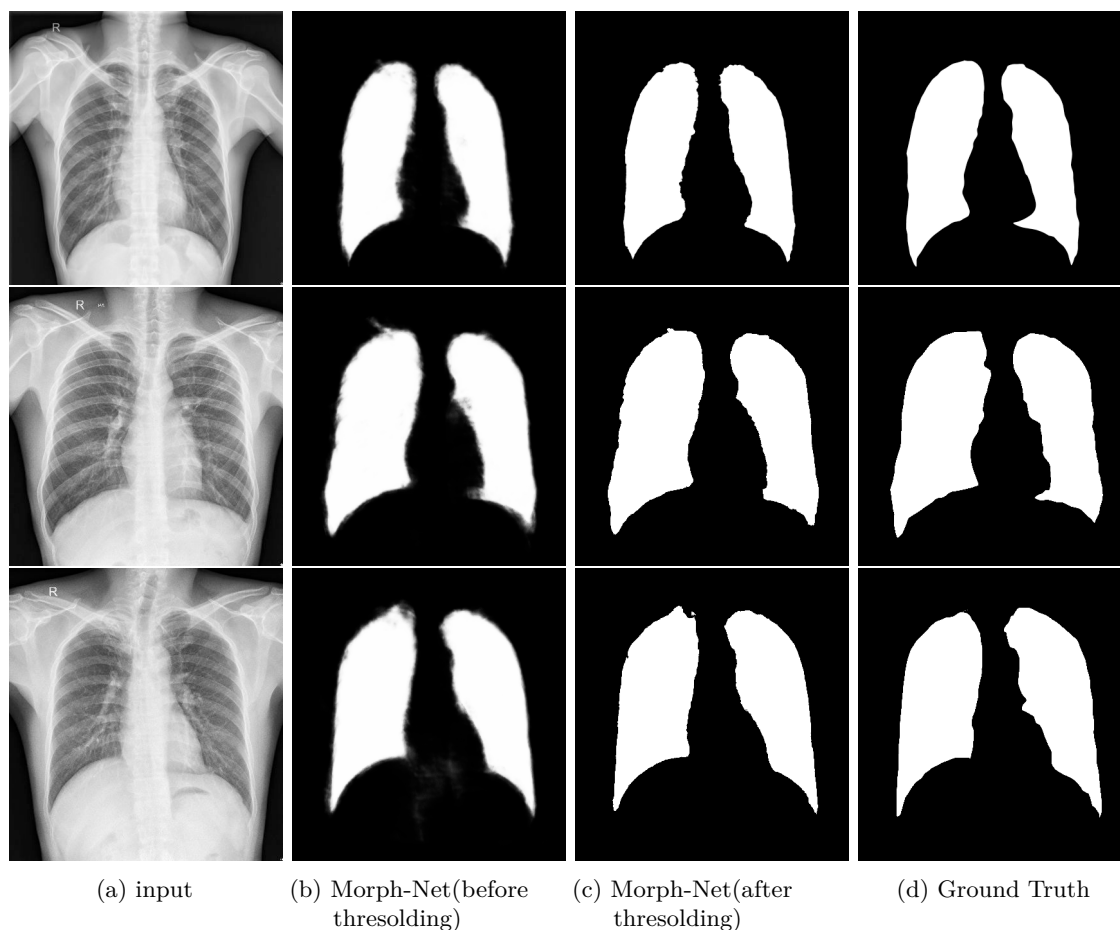
Figure 4.9.: Results obtained by 2D Morph-Net on Shenzhen dataset[34] before the thresholding.

4.3.3. Lung image segmentation

Unlike blood vessel segmentation here we have tried to segment a large blob rather than thin structures. Lung segmentation is an important biomedical image segmentation problem because it serves as the first step towards the diagnosis of various disease including tuberculosis detection. In this experiment. Since the output of the network has large blob, the network should have large receptive field. Hence for lungs segmentation, we have to concentrate upon large spatial information, we have employed a U-Net type network architecture. In Figure 4.8 U-Net architecture using Morphological block (dilation-erosion

followed by linear combination of feature maps) is shown. Note that, here max-pool is used to down sample the image features and 2D up sampling to get back to same size as input. 2D Up Sampling scales up the image by using nearest neighbour. After each 2D morphological block batch normalization is used as regularizer. We observe that batch normalization helps the morphological network in fast optimization and also increases the stability of the network. In the last layer sigmoid activation function is taken to produce the output between $[0, 1]$. Note that we have not taken any other activation function in between.
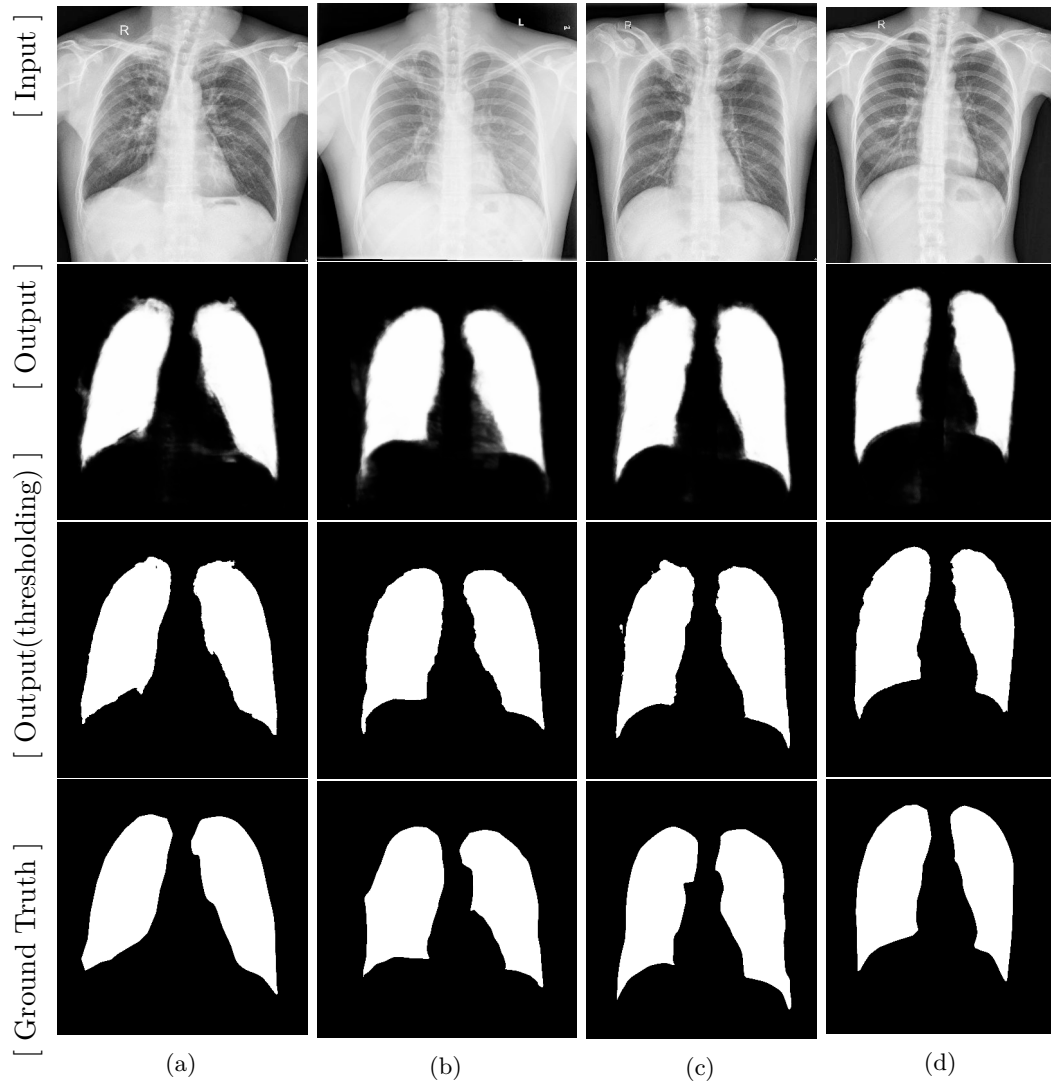
Figure 4.10.: Results on shenzhen dataset with failure cases

Experiments

For evaluating the performance of proposed 2D morphological Network, we have used publicly available Shenzhen chest X-Ray imageset [34]. This dataset is collected by Shenzhen No.3 Hospital in Shenzhen, Guangdong Province, China. The dataset contains 326 normal and 336 X-Ray images showing various manifestations of tuberculosis. Each image is of size approximately 3000×3000. We have randomly selected 90% of the data for training and remaining 10% as test set. All the images are resized to 512×512 before feeding to the network. This problem requires both global and local information, so that a large blob with intricate detail of boundary region (region of interest) can be extracted. This suggests us to employ a U-net [74] type network architecture. The network is trained using binary cross entropy loss and Adam optimizer. The results have been generated with a network trained for 400 epochs with a batch size 2. The output of the network is binarized with a threshold of 0.5 to get the final output. For quantitative evaluation, we have reported the DICE coefficient scores obtained on the test set in Table 4.3. DICE coefficient computes the overlap between the ground-truth and the predicted segmentation mask. So, higher the value, better is the result. In Figure 4.9, we have shown a few results before thresholding. It is seen that the morphological network is able to segment properly as good as the ground truth. Although there are some failure cases as shown in 4.10. In the border places of lungs, the network is producing fuzzy output.
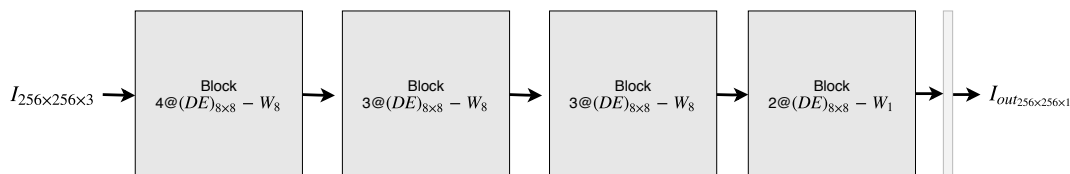
Figure 4.11.: Network Architecture for Document Binarization

4.3.4. Old Document Image Binarization

Binarization and elimination of noise in the historical documents is a challenging task, because of their varying degradation types. Different types of degradation add different types of noise in the binarized image. The task becomes even more difficult when the intensity of the ink colour becomes closer to the intensity of the background colour. Researchers have started working on document image binarzation decades ago, and there are various classical binarization methods available which compute local and global threshold for binarization [61], [78]. Mathematical morphological methods are very useful to filter out different structures or noise in an image. Morphological methods can produce excellent

| (a) Input Image | (b) Morph-Net(Ours) | (c) Ground Truth |

Figure 4.12.: Sample outputs of our proposed binarization approach using ISI-Letters dataset.

results in image processing in which the type filter(s) as well as the shape and size of structuring elements (SEs) or kernels are designed properly. So, we have applied 2D morphological network to address this problem. A simple architecture of 2D morphological network is taken for document image binarization. The network architecture is shown in Figure 4.11. The network has a sequence of 2D morphological block. Each block of this morphological network does some kind of spatial image processing and generates new feature map. Concatenation of multiple such morphological modules with proper training may generate desired feature map in spatial domain. As shown in Figure 4.11, we have

(b) Morph-Net(Ours)

(a) Input Image

(c) Winners Output

Figure 4.13.: Comparison of the outputs of our proposed approach with that of the winner of the ICDAR2017 competition on document image binarization using DIBCO2017 dataset.

(a) Input Image

(b) Morph-Net(Ours)

(c) Winner's Output
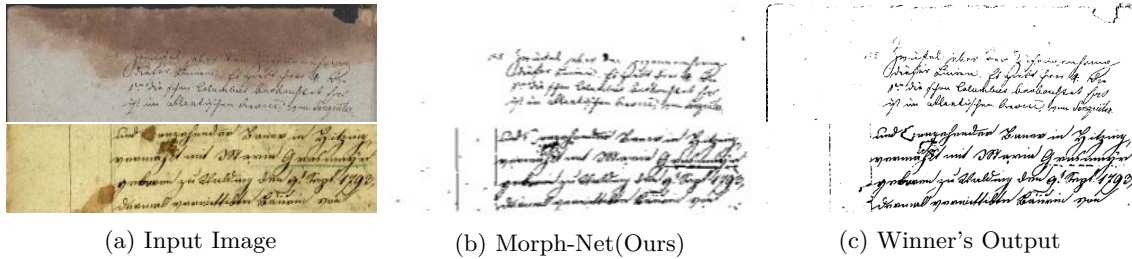
Figure 4.14.: Comparison of the outputs of our proposed method with the winner of the ICFHR2018 competition on document image binarization using H-DIBCO2018 dataset.

taken four such blocks in series, where input image patch is of size $(256 \times 256 \times 3)$ and output is of size $(256 \times 256 \times 1)$. Each pixel of output denotes the probability that the pixel belong to text. Note that in the last layer we have used sigmoid activation function to bound the output in [0,1].
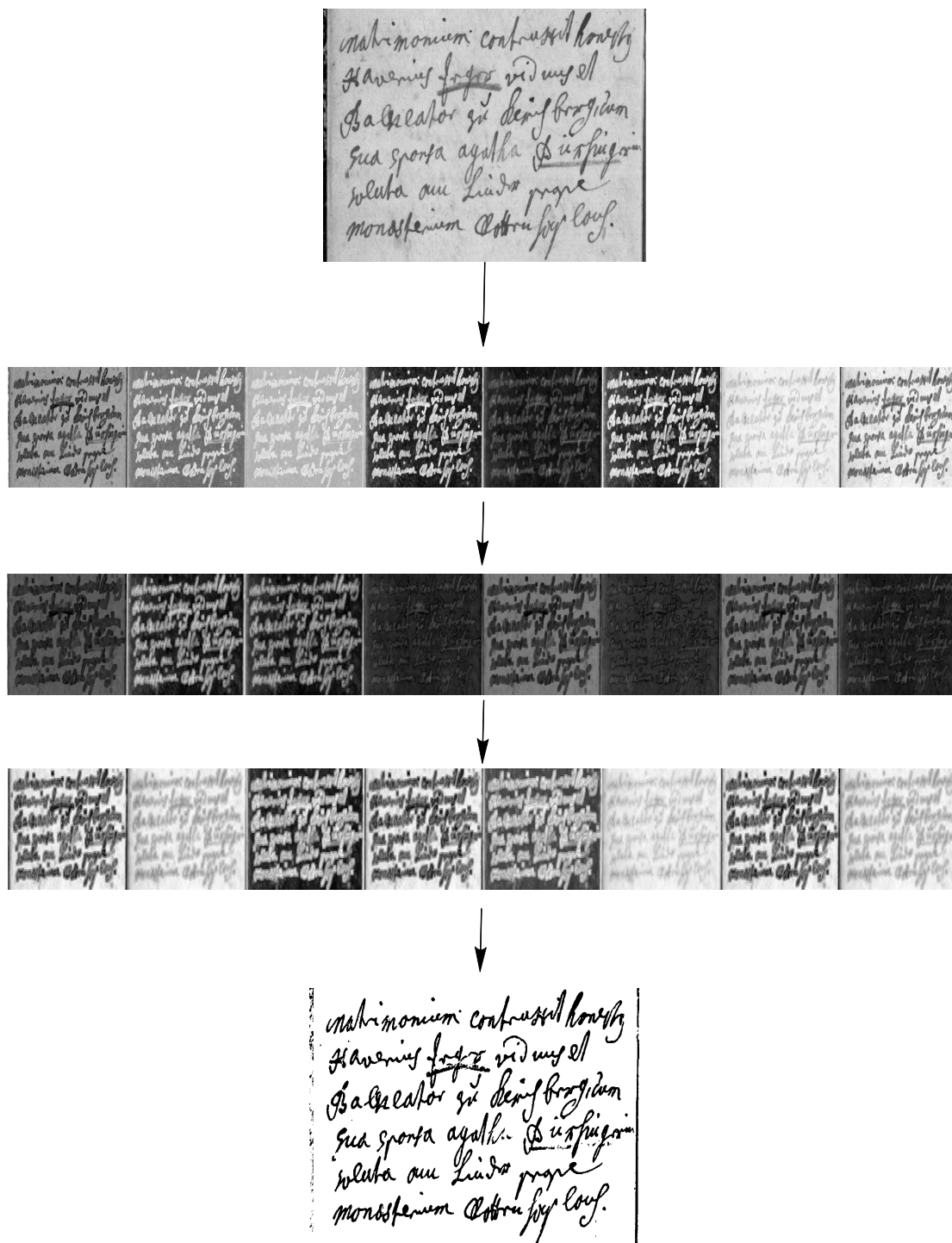
Figure 4.15.: Visualization of different feature maps extracted after each linear combination of dilation and erosion layers on a sample of H-DIBCO 2018 dataset.

Experiments

The proposed 2D morphological network is implemented[3] in python by using library Keras with TensorFlow backend. The experiment is carried out on a machine with Intel Xeon

---

[3] https://github.com/ranjanZ/ICDAR_Binarization

16 core processor and 128GB RAM. We have evaluated our proposed 2D morphological Network for document image binarization on publicly available dataset, namely DIBCO 2017 [64], H-DIBCO 2018 [65] and also on ISI-letters dataset [48]. DIBCO 2017 and H-DIBCO 2018 dataset contain 20 and 10 historical document images respectively. ISI-letter dataset contains 26 letters from the time of second world war that are of different sizes varying from 1499×2343 to 2255×2821. Among the 26 images we have adopted 10 images for our experiment. For the training of the 2D Morphological Network we have used the samples from DIBCO 2009 and H-DIBCO 2016 dataset and also 50% samples of ISI-letters dataset. For data preparation, high resolution document images are split into over lapping patches of size 256×256 with stride 50. The patches are then individually fed into the network for prediction of binarized images. The output image is reconstructed by mosaicing the patches properly. The overlapped areas are averaged to implement superposition. The network optimization settings and evaluation matrices are described. We have provided some sample outputs along with their ground truth images in Figure 4.13 and Figure 4.14 for DIBCO 2017 and H-DIBCO 2018 datasets. From the figures it is clear that our approach is able to binarize the handwritten as well as machine printed documents even if various types of degradation including back impression and page folding marks are present. Moreover, the approach is able to extract the pen strokes properly even if the difference between the foreground and background colour intensity is very low (see Figure 4.13, Figure 4.14). Figure 4.15 shows the step by step intermediate outputs after each convolution layer of the proposed binarization approach using a sample from H-DIBCO 2018 dataset. We have also provided the outputs of our approach on handwritten letters from ISI archive in Figure 4.12.

Note that in association with ICDAR, competitions on old document image binarization were organized in 2017 and 2018 on the DIBCO 2017 and H-DIBCO 2018 datasets. We have compared the output images visually with the output of the winning methods for DIBCO 2017 and H-DIBCO 2018 datasets (Figure 4.13, Figure 4.14). For the outputs of DIBCO 2017 handwritten and machine printed documents (Figure 4.13), the output of our proposed approach is almost equally clear as the output of the winner approach. For H-DIBCO 2018 the Figure 4.14 clearly depicts that the output of our approach is equally good with the winner approach and visually is more noise free than the output of the approach in the second position.

4.3.5. Image Dehazing

Image De-hazing is one of the trending problems of image to image transformation. Haze occurs due to scattering of the light by the particles present in the atmosphere. As a result, only a part of the light that is reflected from the objects, reaches the camera. In addition,

a veil like layer of uniform light is superposed on the image. So, we get a reduced contrast and color shifted image. The hazy image formation process is modeled as [38]

$$\mathbf{I}(\boldsymbol{x}) = t(\boldsymbol{x})\mathbf{J}(\boldsymbol{x}) + (1 - t(\boldsymbol{x}))\mathbf{A}, \tag{4.7}$$

where $\mathbf{J}(\boldsymbol{x})$ and $\mathbf{I}(\boldsymbol{x})$ are the haze-free and the observed hazy image, respectively. $t(\boldsymbol{x}) \in [0, 1]$ is the scene transmittance that controls the haze in a pixel $\boldsymbol{x}$, and $\mathbf{A}$ denotes the uniform environmental illumination. We modify the model by introducing space-variant environmental illumination $\mathbf{A}(\boldsymbol{x})$ as

$$\mathbf{I}(\boldsymbol{x}) = t(\boldsymbol{x})\mathbf{J}(\boldsymbol{x}) + (1 - t(\boldsymbol{x}))\mathbf{A}(\boldsymbol{x}), \tag{4.8}$$

to accommodate more general situations. Note that $\mathbf{A}(\boldsymbol{x})$ varies much slowly than $t(\boldsymbol{x})$.
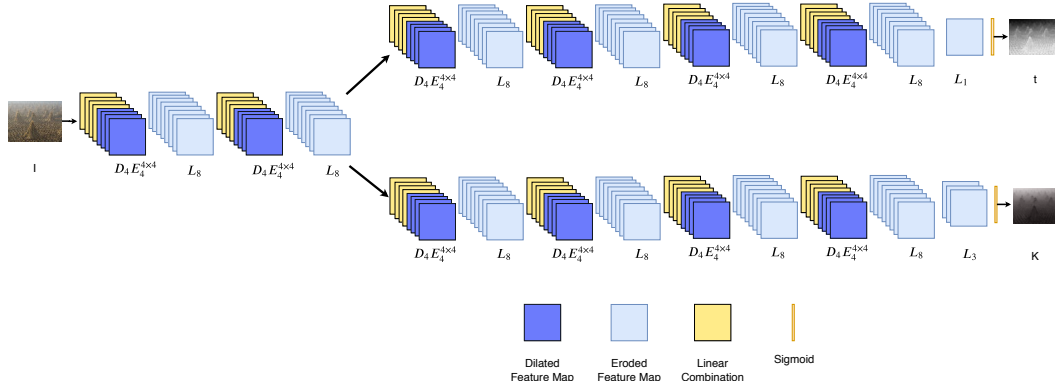
Figure 4.16.: Architecture of the 2D Morph-Net utilized for image dehazing. The network outputs both transmittance and airlight maps which are later utilized to obtain the dehazed image.

Image dehazing methods try to estimate the $\mathbf{J}(\boldsymbol{x})$ having only the $\mathbf{I}(\boldsymbol{x})$ and this usually requires the estimation of $t(\boldsymbol{x})$ and $\mathbf{A}(\boldsymbol{x})$. This makes dehazing an ill-posed problem. Before presenting our method let us re-write equation 4.8 as

$$\mathbf{I}(\boldsymbol{x}) = t(\boldsymbol{x})\mathbf{J}(\boldsymbol{x}) + \mathbf{K}(\boldsymbol{x}). \tag{4.9}$$

Here $(1 - t(\boldsymbol{x}))\mathbf{A}(\boldsymbol{x})$, called *airlight map*, is expressed as $\mathbf{K}(\boldsymbol{x})$. To dehaze an image, we propose a 2D Morph-Net (figure 4.16) that takes hazy image ($\mathbf{I}(\boldsymbol{x})$) as input and produces estimated the airlight map $\tilde{\mathbf{K}}(\boldsymbol{x})$ and the transmittance map $\tilde{t}(\boldsymbol{x})$ as output. As shown in figure 4.16, the 2D morphological network comprises a sequence of 2D morphological blocks. The estimated transmittance map $\tilde{t}(\boldsymbol{x})$ and the airlight map $\tilde{\mathbf{K}}(\boldsymbol{x})$ is computed by two separate paths of the network. This is done so as to capture their dependence while extracting the features specific to both of them. Since, $0 \le t(\boldsymbol{x}) \le 1$ and $0 \le \mathbf{K}(\boldsymbol{x}) \le 1$, we

have used sigmoid activation function in the last layer. From the estimated $\tilde{t}(\boldsymbol{x})$ and $\tilde{\mathbf{K}}(\boldsymbol{x})$ the haze-free image may be obtained as follows.

$$\tilde{\mathbf{J}}(\boldsymbol{x}) = \max\left\{\min\left\{\frac{\mathbf{I}(\boldsymbol{x}) - \tilde{\mathbf{K}}(\boldsymbol{x})}{\tilde{t}(\boldsymbol{x})}, 1\right\}, 0\right\}. \tag{4.10}$$

To train the network, given the hazy image $\mathbf{I}(\boldsymbol{x})$ and corresponding haze-free ground-truth image $\mathbf{J}(\boldsymbol{x})$, we minimize a loss function inspired from the bi-directional consistency loss [55] based on Structural Similarity Index (SSIM) between two images [95]. Specifically, the loss function is given by

$$L = L_1 + L_2. \tag{4.11}$$

L_1 and L_2 are obtained from $\mathbf{I}(\boldsymbol{x})$, $\tilde{\mathbf{I}}(\boldsymbol{x})$, $\mathbf{J}(\boldsymbol{x})$ and $\tilde{\mathbf{J}}(\boldsymbol{x})$ as

$$L_1 = \text{DSSIM}(\mathbf{I}, \tilde{\mathbf{I}}), \tag{4.12}$$

$$L_2 = \text{DSSIM}(\mathbf{J}, \tilde{\mathbf{J}}), \tag{4.13}$$

where

$$\tilde{\mathbf{I}}(\boldsymbol{x}) = \mathbf{J}(\boldsymbol{x})\tilde{t}(\boldsymbol{x}) + \tilde{\mathbf{K}}(\boldsymbol{x}) \tag{4.14}$$

and

$$\text{DSSIM}(\mathbf{X}, \tilde{\mathbf{X}}) = \frac{1}{M}\sum_{i=1}^{M}\frac{1 - \text{SSIM}(P_i, \tilde{P}_i)}{2}. \tag{4.15}$$

P_i and $\tilde{P}_i$ are the ith patches of the $\mathbf{X}$ and $\tilde{\mathbf{X}}$, respectively. M is the total number of patches.

Table 4.4.: Quantitative evaluation on validation dataset of O-HAZE images in terms of SSIM and PSNR metrics.

Method	36.png	37.png	38.png	39.png	40.png	Average
CVPR09[30]	18.18/0.45	16.09/0.49	14.12/0.08	12.88/0.36	14.21/0.39	15.10/0.35
TIP15[107]	17.47/0.50	16.17/0.45	15.14/0.18	14.80/0.41	16.37/0.57	15.99/0.42
TIP16[70]	16.59/0.49	15.76/0.43	13.25/0.19	12.79/0.40	16.53/0.56	14.99/0.41
CVPR16[13]	16.92/0.43	14.98/0.48	15.54/0.34	17.65/0.48	17.04/0.54	16.43/0.45
ICCV17[42]	17.10/0.45	16.47/0.39	16.12/0.12	15.04/0.34	15.95/0.50	16.13/0.36
CVPR18[102]	17.14/0.44	15.29/0.42	14.66/0.11	15.24/0.36	17.78/0.52	16.02/0.36
CVPRW18[103]	**24.67**/0.73	**22.41**/0.66	**23.75**/0.72	**21.91**/0.63	**22.29**/0.68	**23.00**/0.68
Ours	20.22/**0.75**	21.11/**0.75**	19.45/**0.75**	19.63/**0.76**	14.57/**0.64**	19.00/**0.73**

| (a) Input Image | (b) Transmittance | (c) Airlight | (d) Our Output | (e) Ground Truth |

Figure 4.17.: Results of our 2D Morph-Net on three validation image of O-HAZE dataset. Transmittance and airlight map is shown along with ground truth for comparison.

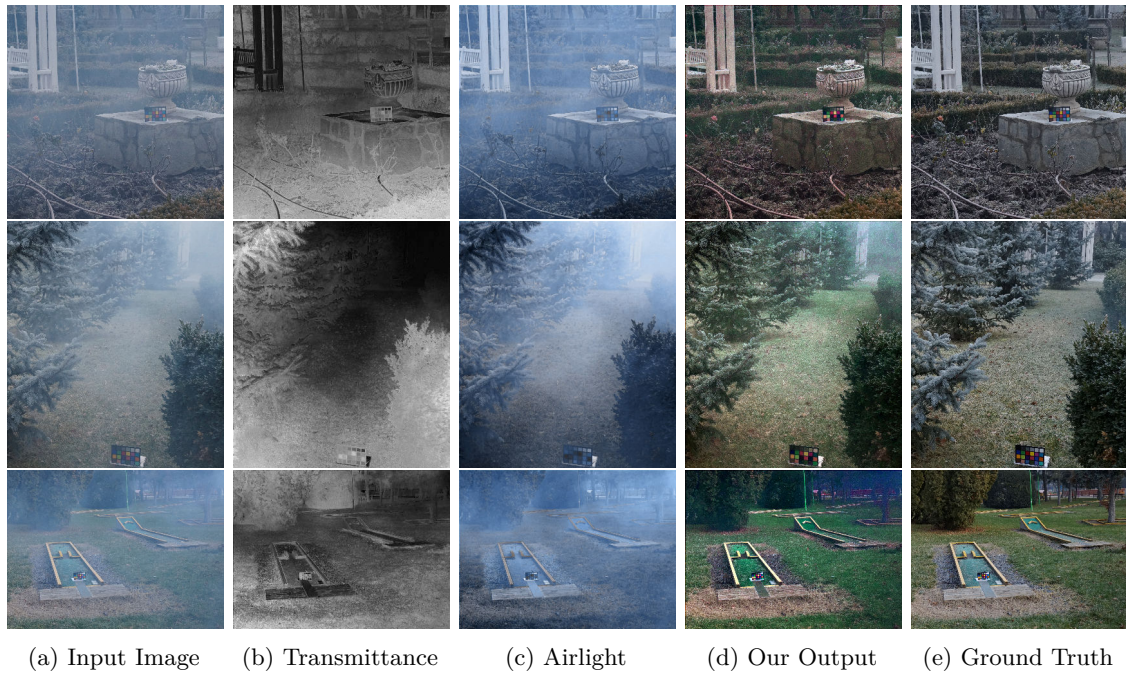| (a) Input Image | (b) Transmittance | (c) Airlight | (d) Our Output | (e) Ground Truth |
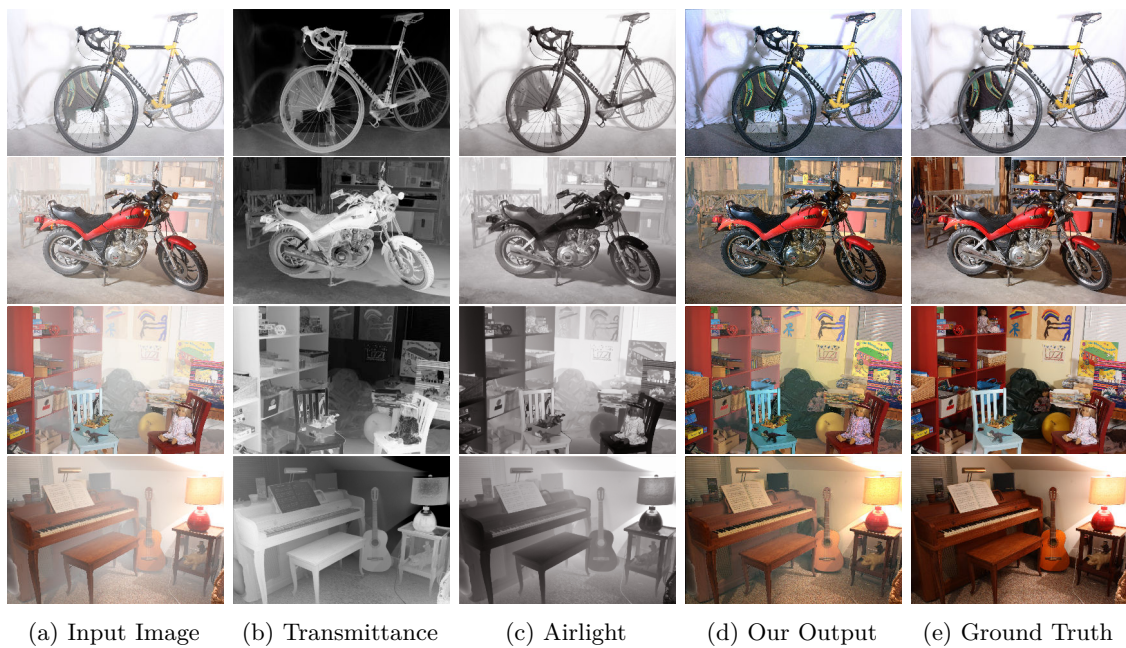
Figure 4.18.: Results of our 2D Morph-Net on Middlebury part of D-Hazy dataset. Transmittance and airlight map is shown along with ground truth for comparison. The network is trained with NYU part of D-Hazy dataset

Table 4.5.: Quantitative results obtained on Middlebury portion of D-Hazy dataset.

Image	He et al. [29] PSNR/SSIM	Ren et al. [69] PSNR/SSIM	Berman [10] PSNR/SSIM	AOD-Net [41] PSNR/SSIM	Morph-Net PSNR/SSIM
Adirondack	16.02/0.82	14.39/**0.89**	**16.74**/0.88	14.18/**0.89**	16.61/0.87
Backpack	14.4/0.85	**16.21**/0.87	12.24/0.82	16.1/**0.91**	14.81/0.85
Bicycle1	12.39/0.81	20.66/0.93	12.61/0.82	**23.21**/**0.96**	16.06/0.89
Cable	**12.95**/**0.7**	7.65/0.64	9.93/0.63	6.95/0.64	11.46/0.66
Classroom1	20.17/0.87	10.91/0.74	**20.95**/**0.89**	10.02/0.72	12.62/0.77
Couch	**18.68**/**0.81**	10.13/0.61	13.76/0.7	10.56/0.63	12.13/0.65
Flowers	**17.73**/0.89	10.47/0.78	17.45/**0.9**	9.25/0.76	14.19/0.83
Jadeplant	**13.48**/**0.69**	7.78/0.6	7.06/0.65	7.65/0.59	12.15/0.63
Mask	**15.88**/0.89	14.15/0.85	14.18/0.84	14.3/**0.91**	15.74/0.86
Motorcycle	13.81/0.79	13.2/0.81	11.6/0.62	12.25/**0.82**	**14.38**/0.81
Piano	**18.66**/**0.86**	12.4/0.71	15.08/0.78	13.89/0.75	17.06/0.80
Pipes	**15.52**/**0.79**	10.9/0.68	13.81/0.74	10.34/0.69	14.92/0.77
Playroom	**17.7**/**0.85**	13.42/0.77	17.64/ 0.83	13.24/0.78	15.71/0.80
Playtable	**18.58**/**0.9**	15.09/0.86	16.63/0.88	14.73/0.86	15.67/0.86
Recycle	12.5/0.82	**18.3**/**0.95**	13.43/0.88	16.62/0.9	13.25/0.88
Shelves	15.47/0.83	**20.43**/**0.94**	16.9/0.88	16.52 0.92	17.03/0.89
Shopvac	**13.87**/**0.8**	7.62/0.66	11.58/ 0.78	6.89/0.64	10.47/0.72
Sticks	16.96/0.9	20.5/**0.96**	20.41/0.93	19.13/**0.96**	**22.10**/**0.96**
Storage	**17.38**/**0.88**	11.23/0.82	16.36/**0.88**	10.24/0.79	17.33/0.87
Sword1	15.06/0.87	**15.48**/**0.91**	12.57/0.83	14.29/**0.91**	14.74/0.87
Sword2	15.66/0.89	12.89/0.88	14.89/0.88	12.8/ **0.9**	**16.06**/0.89
Umbrella	10.4/0.8	**14.92**/0.9	9.63/0.72	14.58/**0.91**	11.61/0.84
Vintage	14.63/0.86	**19.27**/**0.96**	14.09/0.83	16.82/0.94	15.01/0.88
Average	**15.56**/**0.83**	13.82/0.81	14.33/0.81	13.35/0.82	14.83/0.82

Experiments

The network is trained on the training images of O-HAZE dataset [3] until convergence of the training error. We have also trained the 2D Morphological network for dehazing using NYU dataset of D-Hazy dataset and test them with Middlebury portion of D-Hazy dataset. In Table 4.4 and Table 4.5, we have reported the PSNR and SSIM of the results obtained on the validation images of the dataset O-HAZE and Middlebury portion of D-Hazy dataset respectively. We have reported the results with comparison in the Table 4.5. Some qualitative results are shown in figure 4.17 and figure 4.18 for qualitative evaluation and comparison. It is seen that with morphological network is able to remove noise as other methods. We have also shown on real dehazed images in Figure 4.19 which is produced by morph-Net trained on NYU portion of D-Hazy dataset.

(a) Input Image (b) Transmittance (c) Airlight (d) Our Output
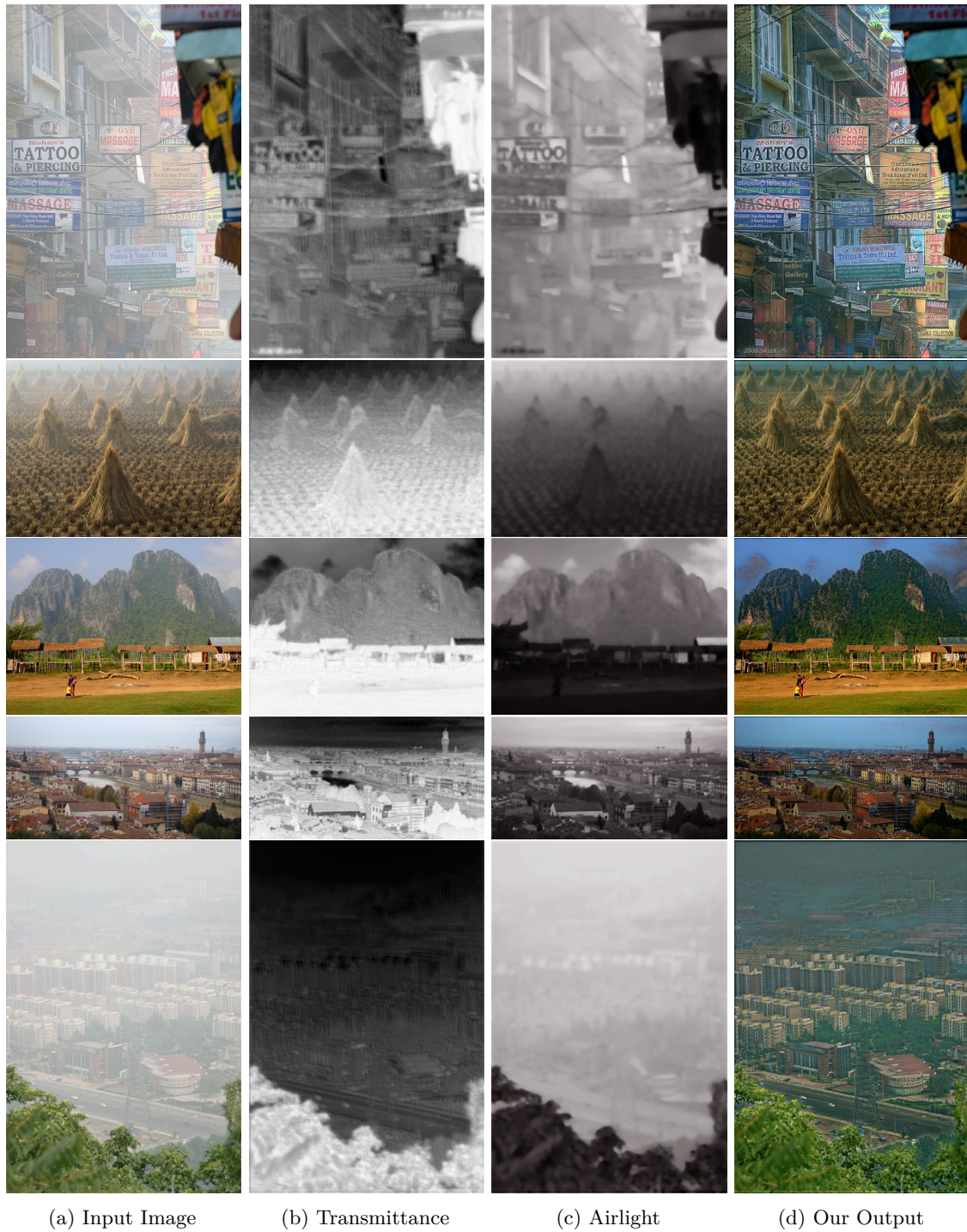
Figure 4.19.: Results of Morphological Network on real outdoor images. Transmittance and airlight is shown along with it.

4.4. Summary

In this chapter we have extended the 1D morphological block to 2D morphological block with all the characteristics of 1D block. We have built 2D morphological network using these 2D morphological blocks. Then we have formed several CNN like networks having appropriate architecture. Efficacy of 2D morphological network is evaluated by applying on various computer vision problems like classification of Aerial image, segmentation of blood vessel in fundus image, segmentation of lung from X-ray images, binarization of old document images and image dehazing. The network performs satisfactorily in those problems, which suggests enough potential to be applied to many more problems.

Part V.

Conclusion

Chapter 5.

Conclusion and Future Scope of Work

Morphological operators are well known for their ability to directly process the shape. However, major hurdles in devising morphological algorithms are deciding the sequence of operations along with designing the appropeiate structuring elements. In this thesis we have proposed a novel network with morphological operations similar to neural network and its variant like convolution neural network. By this way we have tried to solve the problems of designing morphological algorithms. In other words, the proposed network successfully learns the structuring elements, *i.e., 2D dilation and 2D erosion* as well as sequence of morphological operators to be applied. We have defined morphological blocks in real space and put them in a layer form in the proposed Network.

5.1. Conclusion

First we try to see whether structuring elements can be learns using backpropagation algorithm. We propose a morphological network that emulates classical morphological filtering consisting of a series of erosion and dilation operators with trainable structuring elements. The network is successfully applied to filter out (de-emphasize or remove) different structures from an image. For the verification we have considered image de-raining problem. We observe that a single morphological network containing dilation and erosion layers is able to de-rain an image with rain-droplets of arbitrary shape and size. It achieves similar performance with the contemporary CNNs for this task with a fraction of trainable parameters (network size).

Although we are able to learn the structuring elements for gray scale image de-raining task, most of the image restoration tasks are based on removing noise in the color images. We know that morphological opening and closing operations are generally used to remove the noise in the images. Morphological opening and closing are implemented as a sequential combination of dilation and erosion. Although here we use a sequence of dilation and erosion operations, it is difficult to determine in which order opening and closing be used to solve the problem. However, it is observed that when the proposed network is successfully

trained, it may set appropriate order of morphological operations to solve the given problem. So, we have built a network that creates the effect of opening and closing with dilation and erosion layers placed sequentially in two different parallel paths and then by taking a linear combination of their outputs. With the help of the properties of morphological opening and closing operations, we have shown that opening-closing may happen inside the network in an appropriate. To establish the efficacy of the proposed network, we have applied it on two classes of image restoration problems, namely de-raining and de-hazing. Experiment shows that networks with opening and closing layer can handle image restoration tasks. It raises a question- "whether would it be possible to perform classification tasks using dilation and erosion operations like neural network?".

In search of the answer we built dense network using dilation and erosion operators and do rigorous theoretical analysis using min or max operations for feature classification. We have proposed a novel trainable morphological network using dilation and erosion operators as neurons. It is shown that these operators, in conjunction with linear combinations, represent a sum of hinge functions. We call this structure as 1D morphological block. We have proved that by taking two sequential morphological blocks any continuous function can be approximated. However, we are not able to show analytically that a single morphological block can approximate any continuous function. However, we have shown experimentally that a single morphological block is able to perform classification task on different datasets like MNIST, Fashion MNIST, CIFAR-10, SVHN. The performance of the proposed dense morphological block is similar and sometimes better than fully connected neural network. Further we have also explored the different properties of morphological network for multiple layers. After exploring the theoretical aspects of the dense morphological network, It can be readily extended to 2D morphological network forming a CNN like network. More specifically, We have extended 1D morphological network to 2D morphological network using 2D morphological block(s). Efficacy of 2D Morph-Net is evaluated by applying it on various computer vision problems like classification of Aerial image, segmentation of blood vessel in fundus image, segmentation of lung from X-ray images, binarization of old document images and image dehazing.

5.2. Future scope of work

In this work we have tried to generalize the use of morphological dilation and erosion operation in the form of neurons in a network. This opens new direction of research using morphological neurons. Dilation and erosion operation are defined in terms of max and min operations, respectively. As they are piece-wise differentiable, a single component of the structuring element is updated while backpropagation is performed in the network. This single component updating at a time makes training of the structuring element (in turn

the network) very slow. We have used a soft version of dilation and erosion to overcome this problem. A soft dilation used in our work is defined as

$$\boldsymbol{x}\,\hat{\oplus}\,\boldsymbol{s} = \frac{1}{\beta}\log\left(\sum_k e^{(x_k+s_k)\beta}\right), \tag{5.1}$$

We can also use other soft versions as stated below and should check if there is any additional advantage.

$$\boldsymbol{x}\,\hat{\oplus}\,\boldsymbol{s} = (\sum_k (x_k+s_k)^\beta)^{\frac{1}{\beta}} \tag{5.2}$$

$$\boldsymbol{x}\,\hat{\oplus}\,\boldsymbol{s} = \sum_j (x_j+s_j)\frac{e^{(x_j+s_j)\beta}}{\sum_k e^{(x_k+s_k)\beta}} \tag{5.3}$$

$$\boldsymbol{x}\,\hat{\oplus}\,\boldsymbol{s} = \sum_j (x_j+s_j)\frac{(x_j+s_j)^\beta}{\sum_k (x_k+s_k)^\beta} \tag{5.4}$$

From the chapter 3, we know that a single morphological block represents a sum of hinge hyperplanes. If we take multiple such morphological blocks, optimization of such network becomes slow. Convex-concave procedure is popular in optimizing function which are in the form of sum of hinge functions. So, Convex-concave procedure maybe used to optimize each morphological block. Although we have proved that two morphological blocks can approximate any continuous function provided there is no bound in number dilation or erosion neurons. So, it is beneficial to check if there is any bound on number of morphological neurons in each morphological layers, if there is no restriction in number of layers. In this manuscript we have applied morphological networks in some computer vision problems and the network performs satisfactorily on these problems. It exhibits enough potential to be applied to many more problems in image processing domain and one may explore other applications too.

Part VI.

Appendix

Appendix A.

Proofs of Theorems

A.1. Proof of Soft maximum

The soft dilation operation converges to the dilation operation as $\beta \to \infty$. This is true for the soft erosion operation also.

The proof if standard. We include it here for completeness. Let $a_k = x'_k + s_k$. We have

$$
\begin{aligned}
\boldsymbol{x} \oplus \boldsymbol{s} &= \lim_{\beta \to \infty} \log\left(\sum_k e^{a_k\beta}\right)/\beta \\
&= \lim_{\beta \to \infty} \frac{\sum_k a_k e^{a_k\beta}}{\sum_k e^{a_k\beta}} \text{(by L'Hospital's rule)} \\
&= \lim_{\beta \to \infty} \sum_j \frac{a_j e^{a_j\beta}}{\sum_k e^{a_k\beta}} \\
&= \lim_{\beta \to \infty} \sum_j \frac{a_j}{1 + \sum_{k \neq j} e^{(a_k - a_j)\beta}} \\
&= \sum_j \lim_{\beta \to \infty} \frac{a_j}{1 + \sum_{k \neq j} e^{(a_k - a_j)\beta}} \\
&= \max_j a_j.
\end{aligned}
$$

This completes the proof.

A.2. Single morphological block as a sum of hinge functions

Proposition 2. *The function computed by a single layer Morph-Net with n dilation and m erosion neurons followed by a linear combination layer computes $\mathcal{M}(\boldsymbol{x})$, which is a sum of multi-order hinge functions.*

Proof. As defined in the main paper the computed $\mathcal{M}(\boldsymbol{x})$ has the following form.

$$
\mathcal{M}(\boldsymbol{x}) = \sum_{i=1}^{n} \omega_i^+ z_i^+ + \sum_{j=1}^{m} \omega_j^- z_j^-, \tag{A.1}
$$

83

where z_i^+ and z_j^- are the output of ith dilation neuron and j^{th} erosion neuron, respectively and ω_i^+ and ω_j^- are the weights of the the linear combination layer. Replacing the z_i^+ and z_j^- with their expression, the equation becomes the following.

$$\mathcal{M}(\boldsymbol{x}) = \sum_{i=1}^{n} \omega_i^+ \max_{k}\{x_k' + s_{ik}^+\} + \sum_{i=1}^{m} -\omega_i^- \max_{k}\{s_{ik}^- - x_k'\}, \tag{A.2}$$

where s_{ik}^+ and s_{ik}^- denote the kth component of the ith structuring element of dilation and erosion neurons, respectively. The above equation can be further expressed in the following form:

$$\mathcal{M}(\boldsymbol{x}) = \sum_{i=1}^{n} \alpha_i^+ \max_{k}\{\theta i^+ x_k' + \rho_{ik}^+\}$$
$$+ \sum_{i=1}^{m} \alpha_i^- \max_{k}\{\theta_i^- x_k' + \rho_{ik}^-\}, \tag{A.3}$$

where θ_i^+, θ_i^-, ρ_{ik}^+ and ρ_{ik}^- are defined in the following way:

$$\theta_i^+ = \begin{cases} \omega_i^+ & \text{if } \omega_i^+ \geq 0, \\ -\omega_i^+ & \text{if } \omega_i^+ < 0, \end{cases} \qquad \theta_i^- = \begin{cases} -\omega_i^- & \text{if } \omega_i^- \geq 0, \\ \omega_i^- & \text{if } \omega_i^- < 0, \end{cases}$$

$$\rho_{ik}^+ = \begin{cases} s_{ik}^+ \omega_i^+ & \text{if } \omega_i^+ \geq 0, \\ -s_{ik}^+ \omega_i^+ & \text{if } \omega_i^+ < 0, \end{cases} \qquad \rho_{ik}^- = \begin{cases} s_{ik}^- \omega_i^- & \text{if } \omega_i^- \geq 0, \\ -s_{ik}^- \omega_i^- & \text{if } \omega_i^- < 0, \end{cases}$$

$$\alpha_i^+ = \begin{cases} 1 & \text{if } \omega_i^+ \geq 0, \\ -1 & \text{if } \omega_i^+ < 0, \end{cases} \qquad \alpha_i^- = \begin{cases} -1 & \text{if } \omega_i^- \geq 0, \\ 1 & \text{if } \omega_i^- < 0. \end{cases}$$

Now, without any loss of generality, we can write equation A.3 as follows:

$$\mathcal{M}(\boldsymbol{x}) = \sum_{i=1}^{m+n} \alpha_i \max_{k}(\theta_i x_k' + \rho_{ik}), \tag{A.4}$$

where

$$
\theta_i = \begin{cases} \theta i^+ & \text{if } i \leq n \text{ ,} \\ \theta_{i-n}^- & \text{if } n < i \leq m+n \text{ ,} \end{cases}
$$

$$
\rho_{ik} = \begin{cases} \rho_{ik}^+ & \text{if } i \leq n \text{ ,} \\ \rho_{(i-n)k}^- & \text{if } n < i \leq m+n \text{ ,} \end{cases}
$$

$$
\alpha_i = \begin{cases} \alpha_i^+ & \text{if } i \leq n \text{ ,} \\ \alpha_{(i-n)}^- & \text{if } n < i \leq m+n \text{ .} \end{cases}
$$

Clearly, Equation A.4 can be rewritten as

$$
\mathcal{M}(\boldsymbol{x}) = \sum_{i=1}^{l} \alpha_i \phi_i(\boldsymbol{x}), \tag{A.5}
$$

where $l = m + n$, $\alpha_i \in \{1, -1\}$ and

$$
\phi_i(\boldsymbol{x}) = \max_k (\boldsymbol{v}_{ik}^T \boldsymbol{x}' + \rho_{ik}), \tag{A.6}
$$

with

$$
v_{ikt} = \begin{cases} \beta_i & \text{if } t = k, \\ 0 & \text{if } t \neq k. \end{cases} \tag{A.7}
$$

In equation A.6, $\boldsymbol{v}_{ik}^T \boldsymbol{x}' + \rho_{ik}$ is affine and $\alpha_i \phi_i(\boldsymbol{x})$ is a d-order hinge function. Hence $\sum_{i=1}^{l} \alpha_i \phi_i(\boldsymbol{x})$, i.e. $\mathcal{M}(\boldsymbol{x})$, represents a sum of multi-order hinge functions. $\qquad\square$

A.3. Single Morphological block as sum of piecewise linear function

Lemma 5. $\mathcal{M}(\boldsymbol{x})$ *is a continuous piece-wise linear function.*

Proof. From equation A.5, without any loss of generality we can assume that there are t_1 and t_2 number of terms where $\alpha = 1$ and $\alpha = -1$ respectively, then

$$
\mathcal{M}(\boldsymbol{x}) = \sum_{i=1}^{t_1} \phi_i'(\boldsymbol{x}) - \sum_{i=1}^{t_2} \phi_i''(\boldsymbol{x}), \tag{A.8}
$$

where $t_1 + t_2 = l$ and $\phi_i'(\boldsymbol{x}), \phi_i''(\boldsymbol{x})$ are of same form as equation A.6. As sum of PWL functions is also a PWL function, hence each $\sum_{i=1}^{t_1} \phi_i'(\boldsymbol{x})$ and $\sum_{i=1}^{t_2} \phi_i''(\boldsymbol{x})$ are a PWL

function. Now, if $t_1 > 0$, from Proposition 1 we can conclude that $\mathcal{M}(\boldsymbol{x})$ is PWL function since difference of two continuous PWL function is a PWL function . If $t_1 = 0$ then $\mathcal{M}(\boldsymbol{x})$ becomes PWL concave function. Hence, we can say that $\mathcal{M}(\boldsymbol{x})$ is a PWL function. $\square$

It may be noted that if $l < d$ then PWL hyperplane will be in parallel to at least one of the axis. Taking $l \geq d$ results in PWL hyperplane which may span d-dimensional space.

A.4. Two morphological blocks and universal approximation

Here we have proved universal approximation of morphological network, considering bias in the dilation and erosion neurons.

Lemma 6. *Let K be a compact subset of $\mathbb{R}^d$. Then, over K, any hyperplane $w^\top \boldsymbol{x} + b$ can be represented as a linear combination of $(d+1)$ dilation neurons which only depend on K.*

Proof. Since we are in a compact set, there exists $C > 0$ such that $|x_\ell| \leq C$ for any $1 \leq \ell \leq d$. Take

$$s_\ell = -3C\mathbf{1}_{d+1} + 3Ce_{\ell,d+1}, 1 \leq \ell \leq d,$$

where $\mathbf{1}_{d+1}$ is the vector of all ones and $e_{\ell,d+1}$ is the ℓ-th unit vector in $\mathbb{R}^{d+1}$. Then all but the ℓ-th coordinate of s_ℓ are $-3C$, while the ℓ-th coordinate is 0. Then note that, for any $x \in K$, and $1 \leq \ell \leq d$,

$$x_\ell + s_{\ell,\ell} = x_\ell \geq -C > -2C = C - 3C \geq x_j - 3C = x_j + s_{\ell,j},$$

for any $j \neq \ell$. It follows that for any $x \in K$, and $1 \leq \ell \leq d$,

$$\boldsymbol{x} \oplus s_\ell = x_\ell.$$

Note also that the constant 1 can be represented as a dilation neuron by taking $s_{d+1} = -3C\mathbf{1}_{d+1} + (3C+1)e_{d+1,d+1}$. Indeed,

$$1 = s_{d+1,d+1} > -2C = C - 3C \geq x_j - 3C = x_j + s_{d+1,j},$$

for all $1 \leq j \leq d$, and, consequently, for all $x \in K$,

$$1 = \boldsymbol{x} \oplus s_{d+1}.$$

Now given any hyperplane $w^\top \boldsymbol{x} + b$, we can express it exactly as a linear combination of

86

dilation neurons over K:

$$w^\top \boldsymbol{x} + b = \sum_{\ell=1}^{d} w_\ell x_\ell + b = \sum_{\ell=1}^{d} w_\ell (\boldsymbol{x} \oplus s_\ell) + b(\boldsymbol{x} \oplus s_{d+1}).$$

This completes the proof. □

Lemma 7. *Any linear combination of hinge functions $\sum_{i=1}^{m} \alpha_i h^{(k_i)}(\boldsymbol{x})$ can be represented over any compact set K as a two-layer Morph-Net consisting of dilation neurons only.*

Proof. Let $B = \max_{1 \le i \le m} \sup_{\boldsymbol{x} \in K} |h^{(k_i)}(\boldsymbol{x})|$. We now give the architecture of the desired Morph-Net.

1. The first dilation-erosion layer has exactly $(d+1)$ dilation neurons given by $\boldsymbol{x} \oplus s_\ell, 1 \le \ell \le d+1$.

2. The first linear combination layer has $k = \sum_{i=1}^{m}(k_i + 1)$ neurons, with the i-th block of $(k_i + 1)$ neurons outputting the constituent hyperplanes of $h^{(k_i)}(\boldsymbol{x})$. This can be done by Lemma 6.

3. The second dilation-erosion layer just has m dilation neurons, each outputting a hinge function. The ℓ-th neuron is constructed as follows: Write any $\boldsymbol{y} \in \mathbb{R}^k$ as $(\boldsymbol{y}_1^\top, \ldots, \boldsymbol{y}_m^\top)^\top$ where $\boldsymbol{y}_j = (y_{j,1}, \ldots, y_{j,k_j+1})^\top$. We want the output of the ℓ-th neuron to be $\max_{1 \le v \le k_\ell+1} y_{\ell,v}$. So we take $\boldsymbol{t}_\ell = (\boldsymbol{t}_{\ell,1}^\top, \ldots, \boldsymbol{t}_{\ell,m}^\top, -3B)^\top$, where $\boldsymbol{t}_{\ell,j} = -3B\mathbf{1}_{k_j+1}$ for $j \ne \ell$, and $\boldsymbol{t}_{\ell,\ell} = \mathbf{0}_{k_\ell+1}$. Then, for any $j \ne \ell$, $1 \le u \le k_j + 1$, and $1 \le v \le k_\ell + 1$, we have

$$y_{j,u} + t_{\ell,j,u} = y_{j,u} - 3B \le B - 3B = -2B < -B \le y_{\ell,v} = y_{\ell,v} + t_{\ell,\ell,v},$$

and

$$-3B < -B \le y_{\ell,v} = y_{\ell,v} + t_{\ell,\ell,v}.$$

It follows that $\boldsymbol{y} \oplus \boldsymbol{t}_\ell = \max_{1 \le v \le k_\ell+1} y_{\ell,v}$. With this construction, the outputs of the second dilation-erosion layer are the m numbers $h^{(k_i)}(\boldsymbol{x})$.

4. The second linear combination layer just has a single neuron that combines the outputs of the previous layer in the desired way:

$$\boldsymbol{z} \mapsto \sum_{i=1}^{m} \alpha_i z_i.$$

This completes the proof. □

A.5. Properties of multilayer morphological networks

In this section, we record a few properties of multilayer morphological networks.

Lemma 8. *Suppose f and g are real-valued functions on $\mathbb{R}^d$. Then $f = g$ if and only if, for all $r \in \mathbb{R}$, one has equality of the sub-level sets:*

$$f^{-1}(-\infty, r] = g^{-1}(-\infty, r].$$

Proof. The "only if" part is trivial. As for the "if" part, note that we have

$$
\begin{aligned}
f^{-1}\{r\} &= \bigcap_{n \geq 1} f^{-1}(r - 1/n, r] \\
&= \bigcap_{n \geq 1} (f^{-1}(-\infty, r] \setminus f^{-1}(-\infty, r - 1/n]).
\end{aligned}
$$

Same goes for g, and so, by our hypothesis,

$$f^{-1}\{r\} = g^{-1}\{r\} \quad \text{for all } r \in \mathbb{R}.$$

Therefore, for any $x \in \mathbb{R}^d$, we have $x \in g^{-1}\{g(x)\} = f^{-1}\{g(x)\}$, or, in other words, $f(x) = g(x)$. $\qquad\square$

Theorem 4. *The following are true for morphological network architectures.*

(i) *The architecture $D_{m_1}E_0 \to D_{m_2}E_0 \to \cdots \to D_{m_\ell}E_0$ consisting only of dilation layers is equivalent to the architecture $D_{m_\ell}E_0$ with a single dilation layer. A similar statement is true if one considers architectures with only purely erosion layers. Which indicates taking multiple sequential dilation or erosion layer is equivalent of taking a single dilation or erosion layer.*

(ii) *The architecture $D_1E_1 \to D_1$ is not equivalent to D_1E_0. Similarly, it is not equivalent to D_0E_1, and, consequently, the architectures $D_1E_1 \to D_1E_1$ and D_1E_1 are inequivalent.*

(iii) *The architecture $D_1E_1 \to D_1 \to L$ is not equivalent to $D_1E_0 \to L$.*

(iv) *The architecture $D_2E_0 \to D_0E_2 \to D_1$ is not equivalent to $D_2E_0 \to D_1$.*

Proof. (i) Let $x \in R^d$ be the input to the network . Let there be two networks N_1 and N_2. Let there be m_1 and m_2 dilated neurons in, respectively, the first and the second layers of Network N_1. Let the parameters of the network N_1 in the first layer and 2nd layer are $w^1 \in R^{d \times m_1}$ and $w^2 \in R^{l_1 \times l_2}$ respectively. Whereas let there are only single layer with m_1

number of dilated neurons in network N_2 and the parameters are denoted as $u \in R^{d \times m_2}$.
Let $f(x) \in R^{m_2}$ and $g(x) \in R^{m_2}$ are the output from the last layer of network N_1 and N_2
respectively.

For Network N_1

$$y_j = \max_i(x_i + w_{i,j}^1) \quad \forall j \in \{1, 2, ..m_1\} \tag{A.9}$$

$$f_k(x) = \max_j(y_j + w_{j,k}^2) \quad \forall j, k \tag{A.10}$$

For network N_2

$$g_k(x) = \max_j(x_j + u_{j,k}^2) \qquad \forall k, j \tag{A.11}$$

Let

$$S_f^k = \{x \mid f_k(x) \le e_k; e_k \in R\} \tag{A.12}$$

$$S_g^k = \{x \mid f_k(x) \le e_k; e_k \in R\} \tag{A.13}$$

For Network N_1

$$f_k(x) \le e_k; \qquad \forall k \tag{A.14}$$

$$y_i + w_{i,j}^2 \le e_k \qquad \forall k, j \tag{A.15}$$

$$y_i \le e_k - w_{i,j}^2 \qquad \forall k, j \tag{A.16}$$

From Equation A.9 and Equation A.16 we get

$$\max_i(x_i + w_{i,j}^1) \le e_k - w_{i,j}^2 \qquad \forall k, j \tag{A.17}$$

$$x_i + w_{i,j}^1 \le e_k - w_{i,j}^2 \qquad \forall k, j, i \tag{A.18}$$

$$x_i \le e_k - w_{i,j}^2 - w_{i,j}^1 \qquad \forall k, j, i \tag{A.19}$$

Which means

$$x_i \le \min_j(e_k - w_{i,j}^2 - w_{i,j}^1) \qquad \forall k, i \tag{A.20}$$

$$x_i \le e_k - \max_j(w_{i,j}^2 + w_{i,j}^1) \qquad \forall k, i \tag{A.21}$$

$$f_1(x, y) = \max(x + a, y + b)$$

$$f_2(x, y) = \max(f_1 + a_1, g_1 + b_1)$$

$$g_1(x, y) = \min(x + c, y + d)$$
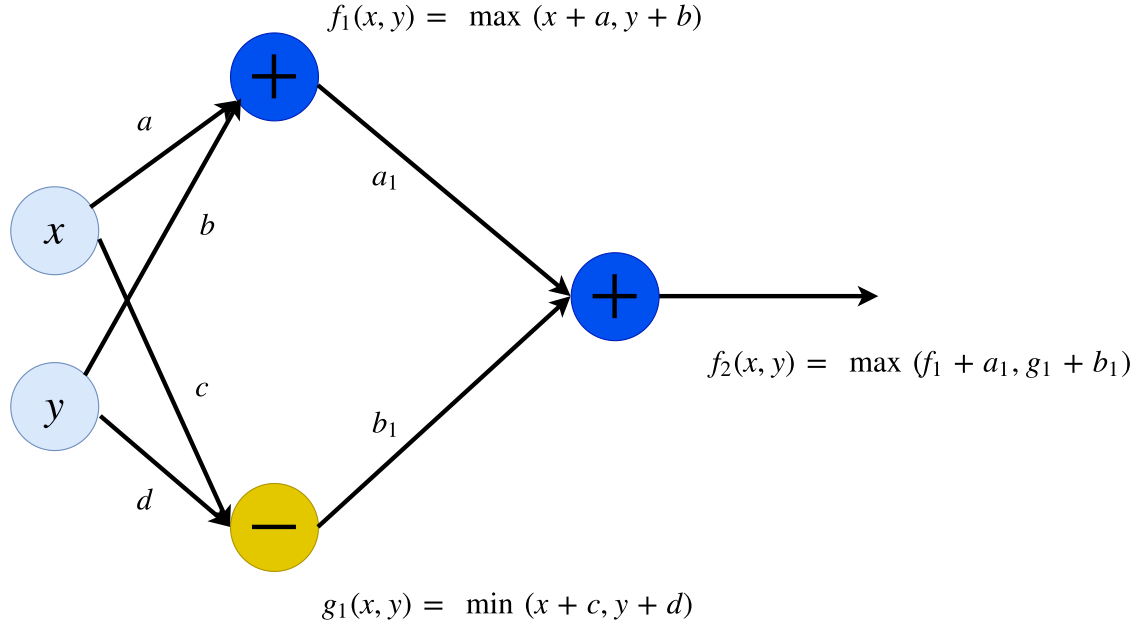
Figure A.1.: A network of architecture $D_1 E_1 \rightarrow D_1$

For network N_2

$$g_k(x) = \max_j(x_j + u_{j,k}^2) \tag{A.22}$$

$$x_i \le (e_k - u_{i,k}) \forall k, i \tag{A.23}$$

To hold the set S_g^k is equal to S_f^k to $\forall k$

$$u_{i,k} = \max_j(w_{i,j}^2 + w_{i,j}^1) \forall i, k \tag{A.24}$$

Hence, from Lemma 8, given a parameter w^1 and w^2 of and 2 layer network N_1, there exist a equivalent single layer network N_2 with dilated neurons u which can represent the same function. From the Equation A.24 we can see the parameters of the single layer network can be constructed considering the longest path from input to output. Recursively we can say it holds for multiple layers. Similar argument can be given in case of erosion layers.

(ii)

For simplicity, we will assume 2-dimensional input. Suppose that the outputs from the first layer are $f_1(x, y)$ and $g_1(x, y)$ where f_1 is the output of a dilation neurones and g_1 is

the output of an erosion neurones (see Figure A.1). We write

$$f_1(x, y) = \max\{x + a, y + b\}, \quad g_1(x, y) = \min\{x + c, y + d\}.$$

After the second layer consisting of a single dilation neurone, we get the output

$$f_2(x, y) = \max\{f_1 + a_1, g_1 + b_1\}.$$

Note that

$$
\begin{aligned}
f_2(x, y) \leq e \iff & f_1 + a_1 \leq e \text{ and } g_1 + b_1 \leq e \\
\iff & f_1 \leq e - a_1 \text{ and } g_1 \leq e - b_1 \\
\iff & (x + a \leq e - a_1 \text{ and } y + b \leq e - a_1) \\
& \text{and } (x + c \leq e - b_1 \text{ or } y + d \leq e - b_1) \\
\iff & (x, y) \in (-\infty, \gamma_1] \times (-\infty, \gamma_2] \\
& \cap ((-\infty, \gamma_3] \times \mathbb{R} \cup \mathbb{R} \times (-\infty, \gamma_4]) \\
\iff & (x, y) \in (-\infty, \gamma_1 \wedge \gamma_3] \times (-\infty, \gamma_2] \\
& \cup (-\infty, \gamma_1] \times (-\infty, \gamma_2 \wedge \gamma_4].
\end{aligned}
$$

Note that $\gamma_1 \leq \gamma_3 \iff a_1 + a \geq b_1 + c$ and $\gamma_2 \leq \gamma_4 \iff a_1 + b \geq b_1 + d$. Therefore, if $a_1 + a \geq b_1 + c$ and $a_1 + b \geq b_1 + d$, then

$$f_2^{-1}(-\infty, e] = (-\infty, \gamma_1] \times (-\infty, \gamma_2].$$

Thus in this case f_2 can be realized in the architecture $D_1 E_0$.

If, however, $a_1 + a < b_1 + c$ and $a_1 + b < b_1 + d$, then

$$f_2^{-1}(-\infty, e] = (-\infty, \gamma_3] \times (-\infty, \gamma_2] \cup (-\infty, \gamma_1] \times (-\infty, \gamma_4],$$

which is not realizable as the sublevel set of a function of $D_1 E_0$ architecture.

(iii) The proof is a simple modification of the proof of (ii). For $\alpha > 0$,

$$
\begin{aligned}
\alpha f_2(x, y) \le e &\iff f_1 + a_1 \le \frac{e}{\alpha} \text{ and } g_1 + b_1 \le \frac{e}{\alpha} \\
&\iff f_1 \le \frac{e}{\alpha} - a_1 \text{ and } g_1 \le \frac{e}{\alpha} - b_1 \\
&\iff (x + a \le \frac{e}{\alpha} - a_1 \text{ and } y + b \le \frac{e}{\alpha} - a_1) \\
&\quad \text{and } (x + c \le \frac{e}{\alpha} - b_1 \text{ or } y + d \le \frac{e}{\alpha} - b_1) \\
&\iff (x, y) \in (-\infty, \gamma_1] \times (-\infty, \gamma_2] \\
&\quad \cap ((-\infty, \gamma_3] \times \mathbb{R} \cup \mathbb{R} \times (-\infty, \gamma_4]) \\
&\iff (x, y) \in (-\infty, \gamma_1 \wedge \gamma_3] \times (-\infty, \gamma_2] \\
&\quad \cup (-\infty, \gamma_1] \times (-\infty, \gamma_2 \wedge \gamma_4].
\end{aligned}
$$

Note that $\gamma_1 \le \gamma_3 \iff a_1 + a \ge b_1 + c$ and $\gamma_2 \le \gamma_4 \iff a_1 + b \ge b_1 + d$.

Therefore, if $a_1 + a \ge b_1 + c$ and $a_1 + b \ge b_1 + d$, then

$$
(\alpha f_2)^{-1}(-\infty, e] = (-\infty, \gamma_1] \times (-\infty, \gamma_2].
$$

Thus, in this case, f_2 can be realized in the architecture $D_1 E_0 \to L$. More explicitly, for $\beta > -1$,

$$
\beta \max\{x + u, y + v\} \le e \iff x \le \frac{e}{\beta} - u \text{ and } y \le \frac{e}{\beta} - v. \tag{A.25}
$$

Equating $\frac{e}{\beta} - u = \frac{e}{\alpha} - a_1 - a$, $\frac{e}{\beta} - v = \frac{e}{\alpha} - a_1 - b$, we can see that one can take $\beta = \alpha, u = a + a_1, v = b + a_1$ to realize the function αf_2 in the $D_1 E_0 \to L$ architecture.

If, however, $a_1 + a < b_1 + c$ and $a_1 + b < b_1 + d$, then

$$
\begin{aligned}
(\alpha f_2)^{-1}(-\infty, e] &= (-\infty, \gamma_3] \times (-\infty, \gamma_2] \\
&\quad \cup (-\infty, \gamma_1] \times (-\infty, \gamma_4],
\end{aligned}
$$

which is not realizable as the sublevel set of a function of $D_1 E_0 \to L$ architecture.

(iv) It can be proved in the same way as (ii) $\qquad \square$

List of Publications Related to the Thesis

[1] Ranjan Mondal, Deepayan Chakraborty, and Bhabatosh Chanda. "Learning 2D Morphological Network for Old Document Image Binarization". In: *2019 International Conference on Document Analysis and Recognition (ICDAR)*. IEEE. 2019, pp. 65–70.

[2] Ranjan Mondal, Moni Shankar Dey, and Bhabatosh Chanda. "Image Restoration by Learning Morphological Opening-Closing Network". In: *Mathematical Morphology-Theory and Applications* 4.1 (2020), pp. 87–107.

[3] Ranjan Mondal, Soumendu Sundar Mukherjee, Sanchayan Santra, and Bhabatosh Chanda. "Morphological Network: How Far Can We Go with Morphological Neurons?" In: *IEEE Transactions on Neural Networks and Learning Systems(under review)* (2019).

[4] Ranjan Mondal, Pulak Purkait, Sanchayan Santra, and Bhabatosh Chanda. "Morphological networks for image de-raining". In: *International Conference on Discrete Geometry for Computer Imagery*. Springer. 2019, pp. 262–275.

List of Other Publications

[1] R. Mondal, S. Santra, and B. Chanda. "Image Dehazing by Joint Estimation of Transmittance and Airlight Using Bi-Directional Consistency Loss Minimized FCN". In: *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. June 2018, pp. 1033–10338. DOI: 10.1109/CVPRW.2018.00137.

[2] Ranjan Mondal and Bhabatosh Chanda. "Anomaly Detection using Context Dependent Optical Flow". In: *Proceedings of the 11th Indian Conference on Computer Vision, Graphics and Image Processing*. 2018, pp. 1–8.

[3] S. Santra, R. Mondal, and B. Chanda. "Learning a Patch Quality Comparator for Single Image Dehazing". In: *IEEE Transactions on Image Processing* 27.9 (Sept. 2018), pp. 4598–4607. ISSN: 1057-7149. DOI: 10.1109/TIP.2018.2841198.

[4] Sanchayan Santra, Ranjan Mondal, Pranoy Panda, Nishant Mohanty, and Shubham Bhuyan. "Image Dehazing via Joint Estimation of Transmittance Map and Environmental Illumination". In: *2017 Ninth International Conference on Advances in Pattern Recognition (ICAPR)*. IEEE. 2017, pp. 1–6.

References

[1] Ricardo de A. Araujo. "A morphological perceptron with gradient-based learning for Brazilian stock market forecasting". In: *Neural Networks* 28 (Apr. 2012), pp. 61–81. ISSN: 0893-6080.

[2] Muhammad Aminul Islam et al. "Deep Morphological Hit-or-Miss Transform Neural Network". In: *arXiv* (2019), arXiv–1912.

[3] Codruta O Ancuti, Cosmin Ancuti, Radu Timofte, and Christophe De Vleeschouwer. "O-HAZE: a dehazing benchmark with real hazy and haze-free outdoor images". In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops.* 2018, pp. 754–762.

[4] Cosmin Ancuti et al. "NTIRE 2018 challenge on image dehazing: Methods and results". In: *CVPR Workshops.* Vol. 1. 2018.

[5] Cosmin Ancuti, Codruta O Ancuti, and Christophe De Vleeschouwer. "D-hazy: A dataset to evaluate quantitatively dehazing algorithms". In: *2016 IEEE International Conference on Image Processing (ICIP).* IEEE. 2016, pp. 2226–2230.

[6] Jesus Angulo. "A mathematical morphology approach to cell shape analysis". In: *Progress in Industrial Mathematics at ECMI 2006.* Springer, 2008, pp. 543–547.

[7] Pierre Baldi, Peter Sadowski, and Daniel Whiteson. "Searching for exotic particles in high-energy physics with deep learning". In: *Nature communications* 5 (2014), p. 4308.

[8] Amy Bearman, Olga Russakovsky, Vittorio Ferrari, and Li Fei-Fei. "What's the point: Semantic segmentation with point supervision". In: *European conference on computer vision.* Springer. 2016, pp. 549–565.

[9] Yoshua Bengio. "Learning deep architectures for AI". In: *Foundations and trends® in Machine Learning* 2.1 (2009), pp. 1–127. ISSN: 1935-8237. DOI: 10.1561/2200000006.

[10] D. Berman, T. Treibitz, and S. Avidan. "Non-local Image Dehazing". In: *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR).* June 2016, pp. 1674–1682. DOI: 10.1109/CVPR.2016.185.

[11] Alan C Bovik. "Handbook of image and video processing". In: (2005).

References

[12]   Leo Breiman. "Hinging hyperplanes for regression, classification, and function approximation". In: *IEEE Transactions on Information Theory* 39.3 (1993), pp. 999–1013.

[13]   Bolun Cai, Xiangmin Xu, Kui Jia, Chunmei Qing, and Dacheng Tao. "Dehazenet: An end-to-end system for single image haze removal". In: *IEEE Transactions on Image Processing* 25.11 (2016), pp. 5187–5198.

[14]   Sema Candemir et al. "Lung segmentation in chest radiographs using anatomical atlases with nonrigid registration". In: *IEEE transactions on medical imaging* 33.2 (2013), pp. 577–590.

[15]   Vasileios Charisopoulos and Petros Maragos. "Morphological perceptrons: geometry and training algorithms". In: *International Symposium on Mathematical Morphology and Its Applications to Signal and Image Processing.* Springer. 2017, pp. 3–15.

[16]   John D Cook. "Basic properties of the soft maximum". In: *UT MD Anderson Cancer Center Department of Biostatistics Working Paper Series; Working Paper 70* (2011).

[17]   Jennifer L. Davidson and Frank Hummer. "Morphology neural networks: An introduction with applications". en. In: *Circuits, Systems and Signal Processing* 12.2 (June 1993), pp. 177–210. ISSN: 1531-5878. DOI: 10.1007/BF01189873.

[18]   Eric Dubois and Anita Pathak. "Reduction of bleed-through in scanned manuscript documents". In: *PICS.* Vol. 1. 2001, pp. 177–180.

[19]   Nick Efford. *Digital image processing: a practical introduction using java (with CD-ROM).* Addison-Wesley Longman Publishing Co., Inc., 2000.

[20]   Raanan Fattal. "Dehazing Using Color-Lines". In: *ACM Trans. Graph.* 34.1 (Dec. 2014), 13:1–13:14. ISSN: 0730-0301. DOI: 10.1145/2651362. URL: http://doi.acm.org/10.1145/2651362.

[21]   Raanan Fattal. "Dehazing using color-lines". In: *ACM transactions on graphics (TOG)* 34.1 (2014), p. 13.

[22]   Richard J Feehs and Gonzalo R Arce. "Multidimensional morphological edge detection". In: *Visual Communications and Image Processing II.* Vol. 845. International Society for Optics and Photonics. 1987, pp. 285–292.

[23]   Gianni Franchi, Amin Fehri, and Angela Yao. "Deep morphological networks". In: *Pattern Recognition* (2020), p. 107246.

[24]   Xueyang Fu, Jiabin Huang, Xinghao Ding, Yinghao Liao, and John Paisley. "Clearing the skies: A deep network architecture for single-image rain removal". In: *IEEE TIP* 26.6 (2017), pp. 2944–2956.

<div align="center">References</div>

[25] Xueyang Fu et al. "Removing rain from single images via a deep detail network". In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.* 2017, pp. 3855–3863.

[26] Ross Girshick. "Fast r-cnn". In: *Proceedings of the IEEE international conference on computer vision.* 2015, pp. 1440–1448.

[27] Xavier Glorot and Yoshua Bengio. "Understanding the difficulty of training deep feed-forward neural networks". In: *Proceedings of the thirteenth international conference on artificial intelligence and statistics.* 2010, pp. 249–256.

[28] Kaiming He, Georgia Gkioxari, Piotr Dollár, and Ross Girshick. "Mask r-cnn". In: *Proceedings of the IEEE international conference on computer vision.* 2017, pp. 2961–2969.

[29] Kaiming He, Jian Sun, and Xiaoou Tang. "Single Image Haze Removal Using Dark Channel Prior". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 33.12 (Dec. 2011), pp. 2341–2353. ISSN: 0162-8828. DOI: `10.1109/TPAMI.2010.168`.

[30] Kaiming He, Jian Sun, and Xiaoou Tang. "Single image haze removal using dark channel prior". In: *IEEE transactions on pattern analysis and machine intelligence* 33.12 (2010), pp. 2341–2353.

[31] HJAM Heijmans. "Morphological Image Operators, Acad". In: *Press, Boston* (1994).

[32] HJAM Heijmans, Michael Buckley, and Hugues Talbot. "Path-based morphological openings". In: *2004 International Conference on Image Processing, 2004. ICIP'04.* Vol. 5. IEEE. 2004, pp. 3085–3088.

[33] A. D. Hoover, V. Kouznetsova, and M. Goldbaum. "Locating blood vessels in retinal images by piecewise threshold probing of a matched filter response". In: *IEEE Transactions on Medical Imaging* 19.3 (2000), pp. 203–210. DOI: `10.1109/42.845178`.

[34] Stefan Jaeger et al. "Two public chest X-ray datasets for computer-aided screening of pulmonary diseases". In: *Quantitative imaging in medicine and surgery* 4.6 (2014), p. 475.

[35] Fuxi Jia, Cunzhao Shi, Kun He, Chunheng Wang, and Baihua Xiao. "Degraded document image binarization using structural symmetry of strokes". In: *Pattern Recognition* 74 (2018), pp. 225–240.

[36] Alexander Kalinovsky and Vassili Kovalev. "Lung image Ssgmentation using deep learning methods and convolutional neural networks". In: (2016).

[37] Diederik P. Kingma and Jimmy Ba. "Adam: A Method for Stochastic Optimization". In: *arXiv:1412.6980 [cs]* (Dec. 2014).

References

[38] Harald Koschmieder. "Theorie der horizontalen Sichtweite". In: *Beitrage zur Physik der freien Atmosphare* (1924), pp. 33–53.

[39] Alex Krizhevsky and Geoffrey Hinton. *Learning multiple layers of features from tiny images*. Tech. rep. University of Toronto, 2009.

[40] Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner. "Gradient-based learning applied to document recognition". In: *Proceedings of the IEEE* 86.11 (1998), pp. 2278–2324.

[41] Boyi Li, Xiulian Peng, Zhangyang Wang, Jizheng Xu, and Dan Feng. "AOD-Net: All-In-One Dehazing Network". In: *Proceedings of the IEEE International Conference on Computer Vision.* 2017, pp. 4770–4778.

[42] Boyi Li, Xiulian Peng, Zhangyang Wang, Jizheng Xu, and Dan Feng. "Aod-net: All-in-one dehazing network". In: *Proceedings of the IEEE International Conference on Computer Vision.* 2017, pp. 4770–4778.

[43] Yu Li, Robby T Tan, Xiaojie Guo, Jiangbo Lu, and Michael S Brown. "Rain streak removal using layer priors". In: *Proceedings of the IEEE conference on computer vision and pattern recognition.* 2016, pp. 2736–2744.

[44] Elena Limonova, Daniil Matveev, Dmitry Nikolaev, and Vladimir V Arlazarov. "Bipolar morphological neural networks: convolution without multiplication". In: *Twelfth International Conference on Machine Vision (ICMV 2019).* Vol. 11433. International Society for Optics and Photonics. 2020, 114333J.

[45] Yu Luo, Yong Xu, and Hui Ji. "Removing rain from a single image via discriminative sparse coding". In: *Proceedings of the IEEE International Conference on Computer Vision.* 2015, pp. 3397–3405.

[46] Diego Marin, Arturo Aquino, Manuel Emilio Gegúndez-Arias, and José Manuel Bravo. "A new supervised method for blood vessel segmentation in retinal images by using gray-level and moment invariants-based features". In: *IEEE Transactions on medical imaging* 30.1 (2010), pp. 146–158.

[47] Jonathan Masci, Jesús Angulo, and Jürgen Schmidhuber. "A learning framework for morphological operators using counter–harmonic mean". In: *International Symposium on Mathematical Morphology and Its Applications to Signal and Image Processing.* Springer. 2013, pp. 329–340.

[48] Bhabatosh Chanda Mayank Wadhwani Debapriya Kundu. "Old Handwritten Document Restoration Based and Clean Text Extraction Using Deep Learning". In: *2nd Workshop on Digital Heritage (WDH' 18) IIIT Hyderabad, Hyderabad* ().

References

[49] Dorra Mellouli, Tarek M Hamdani, Mounir Ben Ayed, and Adel M Alimi. "Morph-CNN: a morphological convolutional neural network for image classification". In: *International Conference on Neural Information Processing*. Springer. 2017, pp. 110–117.

[50] Gili Mendel. "Optical Character Recognition Using Morphological Attributes." In: (1993).

[51] Reza Farrahi Moghaddam and Mohamed Cheriet. "RSLDI: Restoration of single-sided low-quality document images". In: *Pattern Recognition* 42.12 (2009), pp. 3355–3364.

[52] Ranjan Mondal, Deepayan Chakraborty, and Bhabatosh Chanda. "Learning 2D Morphological Network for Old Document Image Binarization". In: *2019 International Conference on Document Analysis and Recognition (ICDAR)*. IEEE. 2019, pp. 65–70.

[53] Ranjan Mondal, Pulak Purkait, Sanchayan Santra, and Bhabatosh Chanda. "Morphological Networks for Image De-raining". In: *International Conference on Discrete Geometry for Computer Imagery*. Springer. 2019, pp. 262–275.

[54] Ranjan Mondal, Sanchayan Santra, and Bhabatosh Chanda. "Dense Morphological Network: An Universal Function Approximator". In: *arXiv preprint arXiv:1901.00109* (2019).

[55] Ranjan Mondal, Sanchayan Santra, and Bhabatosh Chanda. "Image Dehazing by Joint Estimation of Transmittance and Airlight using Bi-Directional Consistency Loss Minimized FCN". In: *CVPR Workshops*. 2018, pp. 920–928.

[56] Susanta Mukhopadhyay and Bhabatosh Chanda. "A multiscale morphological approach to local contrast enhancement". In: *Signal Processing* 80.4 (2000), pp. 685–696.

[57] Yuval Netzer et al. "Reading digits in natural images with unsupervised feature learning". In: *NIPS workshop on deep learning and unsupervised feature learning*. Vol. 2011. 2011, p. 5.

[58] Meindert Niemeijer, Joes Staal, Bram van Ginneken, Marco Loog, and Michael D Abramoff. "Comparative study of retinal vessel segmentation methods on a new publicly available database". In: *Medical imaging 2004: image processing*. Vol. 5370. International Society for Optics and Photonics. 2004, pp. 648–656.

[59] Keiller Nogueira, Jocelyn Chanussot, Mauro Dalla Mura, William Robson Schwartz, and Jefersson A dos Santos. "An Introduction to Deep Morphological Networks". In: *arXiv preprint arXiv:1906.01751* (2019).

[60] Hyeonwoo Noh, Seunghoon Hong, and Bohyung Han. "Learning deconvolution network for semantic segmentation". In: *Proceedings of the IEEE international conference on computer vision*. 2015, pp. 1520–1528.

[61] Nobuyuki Otsu. "A threshold selection method from gray-level histograms". In: *IEEE transactions on systems, man, and cybernetics* 9.1 (1979), pp. 62–66.

[62] Lúcio F. C. Pessoa and Petros Maragos. "Neural networks with hybrid morphological/rank/linear nodes: a unifying framework with applications to handwritten character recognition". In: *Pattern Recognition* 33.6 (June 2000), pp. 945–960. ISSN: 0031-3203. DOI: 10.1016/S0031-3203(99)00157-0.

[63] Bryan A Plummer et al. "Flickr30k entities: Collecting region-to-phrase correspondences for richer image-to-sentence models". In: *Proceedings of the IEEE international conference on computer vision*. 2015, pp. 2641–2649.

[64] Ioannis Pratikakis, Konstantinos Zagoris, George Barlas, and Basilis Gatos. "ICDAR2017 competition on document image binarization (DIBCO 2017)". In: *2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR)*. Vol. 1. IEEE. 2017, pp. 1395–1403.

[65] Ioannis Pratikakis, Konstantinos Zagoris, and Basilis Gatos. "IICFHR 2018 Competition on Handwritten Document Image Binarization (H-DIBCO 2018)". In: *2018 16th International Conference on Frontiers in Handwriting Recognition*. IEEE. 2018.

[66] Kendall Preston Jr and Michael JB Duff. *Modern cellular automata: theory and applications*. Springer Science & Business Media, 2013.

[67] Rabia Rashid, Muhammad Usman Akram, and Taimur Hassan. "Fully Convolutional Neural Network for Lungs Segmentation from Chest X-Rays". In: *International Conference Image Analysis and Recognition*. Springer. 2018, pp. 71–80.

[68] Joseph Redmon and Ali Farhadi. "Yolov3: An incremental improvement". In: *arXiv preprint arXiv:1804.02767* (2018).

[69] Wenqi Ren et al. "Single Image Dehazing via Multi-scale Convolutional Neural Networks". en. In: *Computer Vision - ECCV 2016*. Lecture Notes in Computer Science. Springer, Cham, Oct. 2016, pp. 154–169. DOI: 10.1007/978-3-319-46475-6_10.

[70] Wenqi Ren et al. "Single image dehazing via multi-scale convolutional neural networks". In: *European conference on computer vision*. Springer. 2016, pp. 154–169.

[71] Elisa Ricci and Renzo Perfetti. "Retinal blood vessel segmentation using line operators and support vector classification". In: *IEEE transactions on medical imaging* 26.10 (2007), pp. 1357–1365.

References

[72] G. X. Ritter and P. Sussner. "An introduction to morphological neural networks". In: *Proceedings of 13th International Conference on Pattern Recognition.* Vol. 4. Aug. 1996, 709–717 vol.4. DOI: `10.1109/ICPR.1996.547657`.

[73] David Rivest-Hénault, Reza Farrahi Moghaddam, and Mohamed Cheriet. "A local linear level set method for the binarization of degraded historical document images". In: *International Journal on Document Analysis and Recognition (IJDAR)* 15.2 (2012), pp. 101–124.

[74] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. "U-net: Convolutional networks for biomedical image segmentation". In: *International Conference on Medical image computing and computer-assisted intervention.* Springer. 2015, pp. 234–241.

[75] A Rosenfeld. *AC Kak Digital Picture Processing vol. 2.* 1982.

[76] Azriel Rosenfield and Avinash C Kak. "Digital picture processing". In: *New York: Academic* (1982).

[77] Eyad Haj Said, Diaa Eldin M Nassar, Gamal Fahmy, and Hany H Ammar. "Teeth segmentation in digitized dental X-ray films using mathematical morphology". In: *IEEE transactions on information forensics and security* 1.2 (2006), pp. 178–189.

[78] Jaakko Sauvola and Matti Pietikäinen. "Adaptive document image binarization". In: *Pattern recognition* 33.2 (2000), pp. 225–236.

[79] Dan Schonfeld and John Goutsias. "Optimal morphological pattern restoration from noisy binary images". In: *IEEE Transactions on Pattern Analysis & Machine Intelligence* 1 (1991), pp. 14–29.

[80] J Serra. "Image analysis and mathematical morphology, Acad". In: *Press, NY* (1982).

[81] J Serra. "The Boolean model and random sets". In: *Image Modeling.* Elsevier, 1981, pp. 343–370.

[82] Jean Serra. *Image Analysis and Mathematical Morphology: Vol.: 2: Theoretical Advances.* Academic Press, 1988.

[83] Jean Serra and Pierre Soille. *Mathematical morphology and its applications to image processing.* Vol. 2. Springer Science & Business Media, 2012.

[84] Frank Y Shih. *Image processing and mathematical morphology: fundamentals and applications.* CRC press, 2009.

[85] Frank Y Shih, Yucong Shen, and Xin Zhong. "Development of deep learning framework for mathematical morphology". In: *International Journal of Pattern Recognition and Artificial Intelligence* 33.06 (2019), p. 1954024.

References

[86] João VB Soares, Jorge JG Leandro, Roberto M Cesar, Herbert F Jelinek, and Michael J Cree. "Retinal vessel segmentation using the 2-D Gabor wavelet and supervised classification". In: *IEEE Transactions on medical Imaging* 25.9 (2006), pp. 1214–1222.

[87] Joes Staal, Michael D Abràmoff, Meindert Niemeijer, Max A Viergever, and Bram Van Ginneken. "Ridge-based vessel segmentation in color images of the retina". In: *IEEE transactions on medical imaging* 23.4 (2004), pp. 501–509.

[88] Stanley R Sternberg. "Grayscale morphology". In: *Computer vision, graphics, and image processing* 35.3 (1986), pp. 333–355.

[89] P. Sussner. "Morphological perceptron learning". In: *Proceedings of the 1998 IEEE International Symposium on Intelligent Control (ISIC) held jointly with IEEE International Symposium on Computational Intelligence in Robotics and Automation (CIRA) Intell.* Sept. 1998, pp. 477–482. DOI: 10.1109/ISIC.1998.713708.

[90] Ketan Tang, Jianchao Yang, and Jue Wang. "Investigating haze-relevant features in a learning framework for image dehazing". In: *Proceedings of the IEEE conference on computer vision and pattern recognition.* 2014, pp. 2995–3000.

[91] Luc Vincent. "Morphological area openings and closings for grey-scale images". In: *Shape in Picture.* Springer, 1994, pp. 197–208.

[92] Li Wan, Matthew Zeiler, Sixin Zhang, Yann Le Cun, and Rob Fergus. "Regularization of neural networks using dropconnect". In: *ICML.* 2013, pp. 1058–1066.

[93] Shuning Wang. "General constructive representations for continuous piecewise-linear functions". In: *IEEE Transactions on Circuits and Systems I: Regular Papers* 51.9 (2004), pp. 1889–1896.

[94] Shuning Wang and Xusheng Sun. "Generalization of hinging hyperplanes". In: *IEEE Transactions on Information Theory* 51.12 (2005), pp. 4425–4431.

[95] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. "Image quality assessment: from error visibility to structural similarity". In: *IEEE TIP* 13.4 (2004), pp. 600–612.

[96] Gui-Song Xia et al. "Structural high-resolution satellite image indexing". In: 2010.

[97] Yong Xia, Dagan Feng, and Rongchun Zhao. "Morphology-based multifractal estimation for texture segmentation". In: *IEEE Transactions on Image Processing* 15.3 (2006), pp. 614–623.

[98] Han Xiao, Kashif Rasul, and Roland Vollgraf. "Fashion-mnist: a novel image dataset for benchmarking machine learning algorithms". In: *arXiv preprint arXiv:1708.07747* (2017).

References

[99] Yi Yang and Shawn Newsam. "Bag-of-visual-words and spatial extensions for land-use classification". In: *Proceedings of the 18th SIGSPATIAL international conference on advances in geographic information systems*. 2010, pp. 270–279.

[100] Erik Zamora and Humberto Sossa. "Dendrite morphological neurons trained by stochastic gradient descent". In: *Neurocomputing* 260 (Oct. 2017), pp. 420–431.

[101] Frederic Zana and J-C Klein. "Segmentation of vessel-like patterns using mathematical morphology and curvature evaluation". In: *IEEE transactions on image processing* 10.7 (2001), pp. 1010–1019.

[102] He Zhang and Vishal M Patel. "Densely connected pyramid dehazing network". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2018, pp. 3194–3203.

[103] He Zhang, Vishwanath Sindagi, and Vishal M Patel. "Multi-scale single image dehazing using perceptual pyramid deep network". In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*. 2018, pp. 902–911.

[104] Kai Zhang, Wangmeng Zuo, Yunjin Chen, Deyu Meng, and Lei Zhang. "Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising". In: *IEEE Transactions on Image Processing* 26.7 (2017), pp. 3142–3155.

[105] Kai Zhang, Wangmeng Zuo, and Lei Zhang. "FFDNet: Toward a fast and flexible solution for CNN-based image denoising". In: *IEEE Transactions on Image Processing* 27.9 (2018), pp. 4608–4622.

[106] Yunxiang Zhang, Samy Blusseau, Santiago Velasco-Forero, Isabelle Bloch, and Jesus Angulo. "Max-plus Operators Applied to Filter Selection and Model Pruning in Neural Networks". In: *International Symposium on Mathematical Morphology and Its Applications to Signal and Image Processing*. Springer. 2019, pp. 310–322.

[107] Qingsong Zhu, Jiaming Mai, and Ling Shao. "A fast single image haze removal algorithm using color attenuation prior". In: *IEEE transactions on image processing* 24.11 (2015), pp. 3522–3533.