

# **Automated Grain-Matrix Segmentation in Photomicrographs of Clastic Sedimentary Rocks using Neuro-visual Algorithms**

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

**By**

**Rajdeep Das**

**Under the Supervision of Dr. Kuntal Ghosh, Associate Professor, Machine Intelligence Unit  
and  
Professor Tapan Chakraborty, Geological Studies Unit**



**Machine Intelligence Unit  
Indian Statistical Institute  
Kolkata - 700 108, India**

**May, 2025**

## ACKNOWLEDGEMENT

The completion of this thesis would not have been possible without the expertise, tremendous support and tireless contributions of my PhD supervisors, Dr. Kuntal Ghosh and Prof. Tapan Chakraborty of Indian Statistical Institute, Kolkata.

Special thanks goes to Dr. B. Uma Shankar for his encouraging and knowledgeable suggestions, Prof. Padipta Maji and Prof. Bhabatosh Chanda for their outstandingly supportive and informative course work classes especially those on Advanced Image Processing, and also Late Prof. C. A. Murthy for his elaborative and explanatory course work classes on Pattern Recognition.

Special thanks undeniably goes to Dr. Ajoy Mondol, Debraj Ghosh, Sourav Biswas, Dr. Sujoy Madhab Roy, Dr. Jayanta Kumar Pal, Dr. Tanmoy Basu, Dr. Dipankar Kundu, Shibsankar Roy, Keerthi S Chandran, Joginder Harsolia, Sandipa Roy, Debasis Mazumdar, Roshni Mondal, Anjan Chowdhury, Barnini Bhattacharya especially the last three names mentioned, for their immense help in organizing my thesis in spite of pressure of their own work, and a continually long list of individuals for their continuous support and companionship in preparation of the thesis.

Thanks to Prof. Sanghamitra Bandyopadhyay, Director, ISI, Kolkata for constantly offering such a supportive working environment.

Special thanks to Dr. Bijay Bal for guiding me throughout my research work with his valuable suggestions and continuous motivation.

Thanks to all my seniors, teachers, and office staff of Machine Intelligence Unit, and Center for Soft Computing Research, ISI, Kolkata for their motivating and informative discussions, and help.

I thank everybody in the Geological Studies Unit of the Indian Statistical Institute without the very active cooperation from whom this work would not have been possible. I would especially acknowledge the help received from the faculty and scholars of GSU, particularly Dr. Suchana Taral and Arijit Debnath.

Of course a tremendous portion of credits remain preserved for all my past teachers starting from my school to University for their indispensable knowledge-bearing classes and support in shaping



my higher education journey. Lastly, I would also like to express my sincere gratitude to my parents and family for supporting me throughout my PhD career.

Rajdeep Das

## **Abstract**

Sediments and sedimentary rocks, covering almost sixty-five percent of exposed earth's surface, contain features that allow us to interpret ancient depositional environments, and their role in the evolution of life on earth. Study of sedimentary rocks is also relevant to exploration of geo-economic resources including petroleum, ore minerals, and groundwater as well as research in environmental geology, anthropological studies, and the bi-directional relation of man and environment. One of the important and crucial steps in this study of the detrital sedimentary rocks like sandstone, is to know the size and shape of the grains/clasts (fragments of different minerals or rocks), and the character of the fine-grained matrix (usually clay sized material or chemical precipitates), that surround the individual clasts. Geologists generally use a petrographic microscope to distinguish and study the clasts and the matrix. Traditionally the identified grains/clasts are demonstrated by the geologist on a photomicrograph of the thin section captured on a polarizing microscope. The images contain hundreds of grains/clasts of different sizes and characters. The manual delineation and identification of such grains is extremely time-consuming and tedious. Often the grain boundaries are hazy and almost overlapping, and their exact delineation is highly subjective at places. Therefore, the machine vision and image analysis techniques for grain segmentation from the matrix, and for the measurement of other properties of these grains, appear to be one of the potential tools. Yet, until now, to the best of our knowledge no robust general solution for that has been drawn. Also, for applying state-of-art algorithms in the domain to check their validity in grain-matrix separation is impossible as no generic dataset for this is available at present. In this regard, the present thesis builds a dataset of photomicrographs captured under a petrographic microscope of a thin section of rock samples, and their corresponding ground truths in collaboration with the geologists. This dataset consists of two views, the plane polarized (PPL) view observed under a single polarizer and the cross polarized (XPL) view when the analyzer is inserted as well. Three different approaches in computer vision have been adopted in this thesis to solve the above-mentioned grain-matrix segmentation problem relying upon the concept of mimicking the geologist's eye on the petrographic microscope from various perspectives. The first of these, uses separate spatial filters for three commonly applied modes of vision viz. Vision applying Strong Contrast (VSC), Vision at a Glance (VG), and Vision with Scrutiny (VS). For this, a psychophysics based receptive field model is combined with a clustering algorithm for figure-ground segregation of the XPL images. The second approach employs a Laplacian of

Gaussian filter based mid-level vision model for automated segmentation of the XPL images. It adaptively combines VG and VSC filters mentioned above and ends up with a filter termed as Vision with Adaptive Scrutiny (VAS). The third approach is a Linknet inspired Deep Semantic Grain Segmentation Network that takes cues not only from XPL, but also the PPL image, much like what is often actually done while using the microscope. This is compared with state-of-the-art deep neural networks (itself inspired by visual cortical processing) for image segmentation. The results have been quantitatively analyzed using suitable metrics. The thesis demonstrates that computer vision algorithms, especially the neuro-visually inspired ones, including deep neural nets, have a high potential in automating the difficult problem of grain segmentation from the photomicrographs of clastic sedimentary rocks.

# Contents

|       |                                                                                           |    |
|-------|-------------------------------------------------------------------------------------------|----|
| 1     | Introduction .....                                                                        | 1  |
| 1.1   | A Preliminary Introduction to some Concepts as Mentioned in the title of the Thesis ..... | 1  |
| 1.2   | Motivation and Objectives .....                                                           | 3  |
| 1.2.1 | Motivation behind Automated Grain Segmentation in Sedimentary Rocks .....                 | 3  |
| 1.2.2 | Geological Significance of Grain Size Analysis in Sedimentary Rocks .....                 | 6  |
| 1.2.3 | Objective: Part I- A Novel Sandstone Photomicrograph Dataset Proposed as Benchmark .....  | 11 |
| 1.2.4 | Objective: Part II- Classical Computer Vision based Grain-matrix Segmentation ...         | 11 |
| 1.2.5 | Objective: Part III- Deep Neural Network based Grain-matrix Segmentation .....            | 12 |
| 1.3   | Thesis Organization and Chapter-wise Contributions .....                                  | 13 |
| 2     | Grain-matrix Segmentation: Background and Related Works .....                             | 15 |
| 2.1   | Introduction .....                                                                        | 15 |
| 2.2   | Some Object-background Segmentation Techniques .....                                      | 16 |
| 2.2.1 | Edge-Based Segmentation .....                                                             | 16 |
| 2.2.2 | Energy-Based Segmentation .....                                                           | 19 |
| 2.2.3 | Region-Based Segmentation .....                                                           | 20 |
| 2.2.4 | Threshold-Based Segmentation .....                                                        | 21 |
| 2.2.5 | Visual Perception-Based Segmentation .....                                                | 22 |
| 2.2.6 | Deep Learning-Based Segmentation .....                                                    | 23 |
| 2.3   | Supervised, Unsupervised and self-supervised image segmentation techniques .....          | 25 |
| 2.3.1 | Supervised image segmentation techniques .....                                            | 26 |
| 2.3.2 | Unsupervised image segmentation techniques .....                                          | 26 |
| 2.3.3 | Self-supervised image segmentation techniques .....                                       | 27 |
| 2.4   | Some Concepts in Neuro-visual and psychophysics algorithms used in this thesis .....      | 28 |
| 2.5   | Related Works in Computer Vision based Petrographic Microscopy.....                       | 30 |
| 2.6   | Some metrics used for performance evaluation in image segmentation .....                  | 33 |
| 2.7   | Conclusion .....                                                                          | 34 |

|   |                                                                                                                                               |    |
|---|-----------------------------------------------------------------------------------------------------------------------------------------------|----|
| 3 | Building a Dataset of Clastic Sedimentary Rocks for Segmentation: a Unique and Timely Contribution .....                                      | 35 |
|   | 3.1 Introduction .....                                                                                                                        | 35 |
|   | 3.2 Dataset preparation .....                                                                                                                 | 39 |
|   | 3.3 Methodology for preparing ground truth .....                                                                                              | 42 |
|   | 3.4 Results based on unsupervised segmentation .....                                                                                          | 47 |
|   | 3.5. Conclusion .....                                                                                                                         | 49 |
| 4 | Automatic grain segmentation in cross-polarized photomicrographs of sedimentary rocks using a psychophysics inspired neuro-visual model ..... | 50 |
|   | 4.1 Introduction .....                                                                                                                        | 50 |
|   | 4.1.1 The motivation behind brightness perception modeling for photomicrograph analysis .....                                                 | 52 |
|   | 4.1.2 The psychophysics of brightness-contrast and its modeling .....                                                                         | 53 |
|   | 4.2 Materials and methods .....                                                                                                               | 56 |
|   | 4.2.1 The neural model for brightness perception .....                                                                                        | 57 |
|   | 4.2.1.1 Psychophysical validation of the neural model using brightness induction illusions .....                                              | 60 |
|   | 4.2.2 Filtering the cross-polar photomicrographs in the three proposed vision modes .....                                                     | 65 |
|   | 4.2.3 Clustering the filtered images in CIE Lab space .....                                                                                   | 66 |
|   | 4.2.4 Reducing the number of clusters from twenty to two: choice of threshold using $\beta$ -index .....                                      | 68 |
|   | 4.2.5 Performance measure of final segmentation using Dice score, Accuracy and Jaccard index .....                                            | 70 |
|   | 4.3 Results and discussion .....                                                                                                              | 70 |
|   | 4.4 Conclusion .....                                                                                                                          | 79 |
| 5 | The Mid-level Vision Model of Adaptive Scrutiny for Automated Grain Segmentation in cross-polarized photomicrographs .....                    | 81 |
|   | 5.1 Introduction .....                                                                                                                        | 81 |
|   | 5.2 The Raw Primal Sketch of David Marr .....                                                                                                 | 83 |
|   | 5.3 The Proposed Algorithm for Adaptive Scrutiny .....                                                                                        | 83 |

|       |                                                                                                                                          |     |
|-------|------------------------------------------------------------------------------------------------------------------------------------------|-----|
| 5.4   | Results .....                                                                                                                            | 85  |
| 5.5   | Conclusion .....                                                                                                                         | 87  |
| 6     | Deep Learning Methodology for Automatic Grain-matrix Segmentation in Combined Plane and Cross-polarized Sandstone Photomicrographs ..... | 88  |
| 6.1   | Introduction .....                                                                                                                       | 88  |
| 6.2   | Image Semantic Segmentation using Deep Learning .....                                                                                    | 91  |
| 6.3   | Deep Learning Frameworks For Grain Segmentation .....                                                                                    | 92  |
| 6.3.1 | Network Architecture .....                                                                                                               | 93  |
| 6.3.2 | Loss function comparison .....                                                                                                           | 95  |
| 6.4   | Experiments .....                                                                                                                        | 97  |
| 6.4.1 | Implementation Details .....                                                                                                             | 99  |
| 6.4.2 | Ablation Study on Various Training Sets .....                                                                                            | 101 |
| 6.4.3 | Ablation Study on Loss Function .....                                                                                                    | 103 |
| 6.4.4 | Parameter Comparison .....                                                                                                               | 104 |
| 6.5   | Conclusions .....                                                                                                                        | 105 |
| 7     | Conclusion .....                                                                                                                         | 107 |
| 7.1   | Introduction .....                                                                                                                       | 107 |
| 7.2   | The Geological Significance of the present work .....                                                                                    | 108 |
| 7.3   | Future Direction .....                                                                                                                   | 108 |
|       | Bibliography .....                                                                                                                       | 112 |
|       | Appendix A .....                                                                                                                         | 127 |
|       | Appendix B .....                                                                                                                         | 131 |
|       | Appendix C .....                                                                                                                         | 133 |

# List of Figures

|                                                                                                         |    |
|---------------------------------------------------------------------------------------------------------|----|
| 1.1 Sandstones with rich matrix and little matrix .....                                                 | 6  |
| 1.2 Classification of Sandstone .....                                                                   | 7  |
| 1.3 a. Thin section photomicrograph of a well sorted, well-rounded fine-grained Aeolian sediments       |    |
| b. Photomicrograph of poorly sorted coarser sediments of fluvial deposit .....                          | 9  |
| 1.4 a. SEM photograph of clay mineral                                                                   |    |
| b. Mature sandstone made of quartz                                                                      |    |
| c. Reduction in the sandstone porosity                                                                  |    |
| d. Reduced porosity of the sandstone with precipitation .....                                           | 10 |
| 1.5 Overall organization of the thesis .....                                                            | 13 |
| 2.1 Edge detection based grain segmentation using various edge detection techniques .....               | 18 |
| 2.2 Edge detection method may not give yield the desired segmentation as shown in the two               |    |
| ambiguous regions .....                                                                                 | 19 |
| 2.3 Energy minimization based grain segmentation .....                                                  | 19 |
| 2.4 Segmentation has been through region competition (Jungmann et al., 2014). Classification has        |    |
| been performed into quartz (white) and anhydrite (black) .....                                          | 21 |
| 2.5 Visual comparison among different methods on quartz sandstones .....                                | 23 |
| 2.6 The Linknet Architecture .....                                                                      | 24 |
| 2.7 The U-Net Architecture .....                                                                        | 25 |
| 2.8 A typical sedimentary rock photomicrograph and its ground truth marked by the expert .....          | 28 |
| 2.9 Examples of (a) brightness contrast and (b) brightness assimilation [De Weert & Spillman]           |    |
| .....                                                                                                   | 29 |
| 2.10 The parallel visual channels from the eye to the striate cortex/primary visual cortex (V1). [Image |    |
| courtesy: <a href="https://webvision.med.utah.edu/">https://webvision.med.utah.edu/</a> ] .....         | 30 |
| 3.1 Outline and parts of a petrographic microscope .....                                                | 40 |
| 3.2 (a) PPL image (b) XPL Image (c) GT image .....                                                      | 41 |
| 3.3 The steps followed in preparing the ground truth image .....                                        | 42 |
| 3.4 A sample photomicrograph from our dataset showing the XPL, PPL views and the GT image               |    |
| containing many grains .....                                                                            | 43 |
| 3.5 A sample photomicrograph from our dataset containing comparatively larger grains, and a             |    |
| considerable amount of matrix area unlike most of the images found in literature .....                  | 44 |
| 3.6 Another sample image from our dataset containing larger grains and ambiguous border regions         |    |
| .....                                                                                                   | 45 |
| 3.7 A sample image from our dataset containing a large number of small sized grains .....               | 46 |

|                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       |    |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----|
| 3.8 A sample image from our dataset containing a large number of very small grains. It can be seen that in this image, apart from black and white there are some gray portions in the GT image demonstrating those regions that cannot be marked either as grain or matrix and are hence identified as ambiguous regions in gray color .....                                                                                                                                                          | 47 |
| 3.9 Output images under unsupervised method. (a) original cross-polar image, (b) Otsu's segmentation, (c) Kapur's entropy segmentation, (d) NCut segmentation, (e) LBP segmentation .....                                                                                                                                                                                                                                                                                                             | 48 |
| 4.1 Typical cross-polar photomicrographs from sedimentary rocks displaying the large rock fragments, viz. grains, with the finer-grained sedimentary material, such as clay or silt termed as matrix, occurring in the interstices of the larger grains .....                                                                                                                                                                                                                                         | 53 |
| 4.2 A typical sedimentary rock photomicrograph and its ground truth marked by the expert .....                                                                                                                                                                                                                                                                                                                                                                                                        | 54 |
| 4.3 Examples of (a) brightness contrast and (b) brightness assimilation .....                                                                                                                                                                                                                                                                                                                                                                                                                         | 54 |
| 4.4 Frequency response of the proposed P and M filters .....                                                                                                                                                                                                                                                                                                                                                                                                                                          | 59 |
| 4.5 (a) The Checkerboard stimulus at 8 pixels width per column (b) Horizontal intensity profile through the two test patches of the output image filtered by P channel, demonstrating brightness assimilation from the surrounding, with the left patch appearing brighter .....                                                                                                                                                                                                                      | 61 |
| 4.6 (a) The Checkerboard stimulus at 25 pixels checker dimension (b) Horizontal intensity profile through the two test patches of the output image spatially filtered by P channel demonstrating brightness contrast with respect to surround, with the left patch appearing darker .....                                                                                                                                                                                                             | 63 |
| 4.7 (a) The checkerboard stimulus (b) The corresponding intensity profile of the output image spatially filtered through the proposed M channel displays the sharp undershoots and overshoots in the vicinity of the test patches .....                                                                                                                                                                                                                                                               | 64 |
| 4.8 (a) The classical example of brightness-contrast. (b) The White effect illusion. (c) Output of P channel over (a). (d) Output of M channel over (b) .....                                                                                                                                                                                                                                                                                                                                         | 65 |
| 4.9 A sample cross-polar photomicrograph filtered in the three proposed vision modes. (a) The original image. (b) The "vision-at-a-glance" output (c) The "vision-in-strong-contrast" mode (d) The "vision-with-scrutiny".....                                                                                                                                                                                                                                                                        | 66 |
| 4.10 Results on a typical photomicrograph displaying its VG, VSC, and VS modes .....                                                                                                                                                                                                                                                                                                                                                                                                                  | 72 |
| 4.11 Output images for applying on a lower-resolution image shown in Fig. 4.9 algorithm 4.2 .....                                                                                                                                                                                                                                                                                                                                                                                                     | 73 |
| 4.12 Results on another typical cross-polar photomicrograph (Fig 4.12a) with larger grains in the image that has been captured at higher resolution displaying its VG, VSC, and VS modes (Fig 4.12b, 4.12d, 4.12f) and their corresponding plots of cluster number (after sorting the clusters in ascending order of their centroid norms) vs. $\beta$ index plot (Fig. 4.12c, 4.12e, 4.12g), for binary segmentation of the image after clustering into 20 clusters and optimizing for 20 runs ..... | 75 |



|                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       |     |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| 4.13 Results for different modes on cross-polar image in 4.12a .....                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  | 77  |
| 5.1 Comparison of the automated channel selection method VAS with those in the previous Chapter in terms of Dice scores .....                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         | 86  |
| 5.2 a. Original cross polar image, b. Segmented image by non-automatic visual channel method, c. Segmented image by the currently proposed LOG-based automatic method .....                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           | 86  |
| 6.1 Visually illustrates an overview of a deep learning model for semantic segmentation tasks .....                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   | 93  |
| 6.2 Shows the encoder-decoder architecture in the Lin[LinkNet Module → Encoder-decoder module]kNet [Garcia et al., 2020] module .....                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 | 94  |
| 6.3 Shows the details of convolution modules in the encoder-block (i) .....                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           | 94  |
| 6.4 Shows the details of convolution modules in the decoder-block (i) .....                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           | 95  |
| 6.5 Illustrates some of the images from our dataset, used in this current work, with ground truth segmentation .....                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  | 97  |
| 6.6 Shows the generated average image from some pair of images (cross-polar and plane-polar), used in this current work, from our dataset .....                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       | 98  |
| 6.7 Comparison of various state of the art deep learning frameworks on our dataset .....                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              | 102 |
| 6.8 Illustrates qualitative results obtained using various segmentation architectures .....                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           | 105 |
| 7.1 Thin rock section (50X magnification) under the lab-built polarizing microscope in all the modes, viz. normal PPL (top-left), normal XPL (top-right), attenuated (with white paper) XPL (left bottom), and attenuated (with transparent cellophane paper) XPL (right bottom). It can be clearly seen that only under the attenuated pseudo-color cross polarized mode using transparent cellophane sheets that a crackling has been identified unlike the other modes. The generated image in this mode is also a highly enhanced ideal version for image segmentation .....                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      | 109 |
| B.1 Confusion matrices for the unsupervised algorithms' (stated in this thesis) results on one typical cross-polar image from our built dataset (here 'n' denotes the total number of pixels in the input image and TP, FP, FN, TN denote the number of true positive, false positive, false negative and true negative pixels respectively), for: (a) our proposed psychophysics inspired neuro-visual model in Chapter 4 [Das et al., 2021], (b) our proposed mid-level vision model of adaptive scrutiny, in Chapter 5 [Das and Ghosh, 2022], with used gaussian variance=4, (c) Normalized Cut (NCut) [Shi and Malik, 1997] method, (d) Local Binary pattern (LBP) [Pietikäinen, 2010] method, (e) Otsu [Otsu et al., 1975] thresholding method, (f) Kapur's entropy [Kapur et al., 1985] method of thresholding, (g) Yen's entropy [Yen et al., 1995] method of thresholding, (h) Rosenfeld [Rosenfeld and Torre, 1983] thresholding method, (i) Li's entropy [Li and Lee, 1993] method of thresholding, (j) Abutaleb [Abutaleb, 1989] thresholding method ..... | 132 |

|                                                                                                                                                                                                                                                         |            |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------|
| C.1 Output images under self-supervised method in [Lin et al, 2023] and ground truths for an input xp image. (a) Original xp mode input image, (b) Output image at=0.0095, rt=0.00, (c) Output image at=0.004, rt=0.00 and (d) Ground truth image ..... | <b>134</b> |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------|

## List of Tables

|                                                                                                                                                                                                                                                                                                                                           |            |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------|
| 3.1 List of related type data works. NPA: Not Publicly Available.....                                                                                                                                                                                                                                                                     | <b>38</b>  |
| 3.2 Performance comparison of various unsupervised methods.....                                                                                                                                                                                                                                                                           | <b>48</b>  |
| 4.1 Spatial filters representing the two (M & P) parallel visual channels.....                                                                                                                                                                                                                                                            | <b>58</b>  |
| 4.2 Segmentation indices corresponding to Fig 4.11a for the M, P, and PM filters and comparison with other methods .....                                                                                                                                                                                                                  | <b>74</b>  |
| 4.3 Segmentation indices corresponding to Fig 4.13a for the M, P, and PM filters and comparison with other methods.....                                                                                                                                                                                                                   | <b>78</b>  |
| 5.1 Average performance of some of the well-known existing segmentation techniques for 16 cross-polar images other than those already compared to the proposed method in Chapter 4, in terms of dice score, accuracy, recall, precision, and jaccard index. The figures in bold indicate the best results for each of these metrics ..... | <b>82</b>  |
| 5.2 Table showing the Dice score results for our dataset with Gaussian variance varying from 2 - 18 .....                                                                                                                                                                                                                                 | <b>85</b>  |
| 6.1 Comparing various loss functions, where $w_l = \frac{Nobj}{N}$ and $w_o = \frac{Nbg}{N}$ with $Nobj$ , $Nbg$ and $N$ are the number of pixels within the grain region, number of pixels within the background region and the total number of pixels in an image of the training set, respectively.....                                | <b>96</b>  |
| 6.2 Performance comparison among state of the art segmentation networks on Test-Set2.....                                                                                                                                                                                                                                                 | <b>103</b> |
| 6.3 Compares various architectures concerning parameters.....                                                                                                                                                                                                                                                                             | <b>104</b> |
| A.1 Detailed results of applying the psychophysics inspired neuro-visual model based algorithm for grain matrix segmentation as proposed in chapter 4, with highest values highlighted in boldface .....                                                                                                                                  | <b>127</b> |
| C.1 Average performance comparison of a self supervised method [Lin et al, 2022] recently used with two variable parameters set of values for xp, pp and combined corresponding to 16 triplets in our built dataset.....                                                                                                                  | <b>133</b> |

# Chapter 1

## Introduction

### 1.1 A Preliminary Introduction to some Concepts as Mentioned in the title of the Thesis

**Clastic Sedimentary rocks:** Sedimentary rocks are created by the accumulation and consolidation of sediments, which comprises pieces of pre-existing rocks, minerals, or organic matter. These rocks are transported and deposited by wind, water, or ice. Sediments and sedimentary rocks cover almost sixty-five percent of exposed Earth's surface [Blatt, 1992]. There are two major types of sedimentary rocks – the clastic and non-clastic. They are deposited layer by layer and form what is usually referred to as bedding. The detritus formed through weathering of pre-existing rock material are entrained, transported and deposited by glaciers, rivers, and wind or ocean waves to form clastic sedimentary rocks. Whereas, the rocks that form in situ through the chemical precipitation of material (like most of the limestones or banded iron formations) are known as non-clastic rocks. Among the sedimentary rocks, sandstone is particularly important in understanding the source region (provenance), surface processes, paleogeographic scenario and climate that existed on this planet many million years ago.

**Importance of sedimentary rocks:** An in-depth comprehension of the distribution and attributes of grains and matrix inside sandstone yields useful insights about their processes and environment of deposition, subsequent alteration processes, and potential reservoir properties [Nichols, 2009] which in turn yield information about many aspects of past condition of earth

climate and surface processes including evolution of life forms in the geologic past [Altermann, 2002; Curtis, 1990; Reading, 2009].

**Petrographic analysis:** Petrographic analysis [Tucker, 2001], particularly that of sandstone, is a crucial method of study. It involves the examination of thin sections of rock samples under a polarising microscope [Török et al., 1998]. The study includes examination of texture and mineral identification by observing the optical properties of minerals [Nesse, 2000] such as color, birefringence [Rai and Hanson, 1986], cleavage under polarized light, diagenetic features such as cementation, compaction, dissolution, and replacement. One of the important and crucial steps in this study of the detrital sedimentary rocks is to know the geometry and shape of the clasts (fragments of different minerals or rocks), also termed as grains, and the character of the fine-grained matrix (fine-grained, usually clay sized material or carbonates, that surround the individual sand-sized clasts).

**Requirement of automated grain-matrix segmentation algorithm:** The semi-automated machines (point counters) in the polarising microscope allow the thin section to be moved through a fixed interval measuring a fraction of a millimeter, and the geologist keeps on noting at a particular step (the points) what kind of matrix and grains are there. Scanning the entire slide gives one a quantitative estimation of the proportion of grains and matrix, and also about the type of grains present. Thus, assessing the proportion of framework grains and matrix present in a particular sandstone body through this method would probably require one whole day or more. The automated recognition of grains and matrix will greatly facilitate the petrographic analysis of the sandstone, in terms of the time and energy spent by a geologist. Once the framework grains and matrix can be separated by any suitable image segmentation algorithm, it would be easier to measure other parameters of the grains including their size distribution and shape characteristics in a thin section through automated image analysis.

**Neuro-visual algorithms for automated image segmentation:** The state of the art techniques in performing automated image segmentation and also several other similar tasks in computer vision involve the domain of computational neuroscience. This, in turn, deals with various types of Artificial Neural Networks (ANN) including Deep Neural Networks (DNN). Starting from the works of David Marr [Marr, 1982], many of these algorithms are inspired by the structure and

function of the visual cortex in the mammalian/primate brain. This thesis successfully explores several such models motivated by visual psychophysics (a physics based approach to visual perception) and visual neuroscience, as also many established and recent methods from the Computer Vision (CV) and Machine Learning (ML) domain, in performing this task of automated grain-matrix segmentation.

Having introduced and also briefly explained some of the basic concepts mentioned, either explicitly or implicitly, in the title of the thesis, we now proceed to elaborate upon the motivation and objectives of this thesis in greater detail.

## **1.2 Motivation and Objectives**

In this section, we first attempt to elaborate on the motivation that concerns the present problem of identification and segregation of grain and matrix in sedimentary rocks, its geological significance and the needs for image analysis. In the last three subsections of this, we also elaborate upon the three objectives that have been fulfilled through this research. These three objectives and their realization basically constitutes the three contributory parts of this thesis.

### *1.2.1 Motivation behind Automated Grain Segmentation in Sedimentary Rocks*

Sediments and sedimentary rocks cover almost sixty-five percent of exposed Earth's surface [Blatt, 1992]. Once deposited, the sediments undergo burial by other sediments, resulting in increased pressure and temperature. Under this increased pressure-temperature regime diagenetic processes start by precipitating newer minerals from circulating fluids and the loose sediment is lithified into solid rock. These buried rock bodies again come up to the surface of the earth (i.e.,

becomes ‘exposed’) due to some upheavals related to the large-scale tectonic movement in the earth’s crust.

Sedimentary rocks contain features that allow us to interpret ancient depositional environments, including the evolution of organisms and the environments in which these organisms lived in. These rocks are critically indicative of the surface process and is a source of major clues for assessing the climate condition prevailing in the past [Ruddiman, 2008]. The evidence from sedimentary rocks and its contents further provides information regarding various aspects of earth history that ranges from the evolution of life to the development of the mountain ranges. For example, analyzing the sedimentary rocks containing the fossil records, researchers reconstruct the life forms in geological past (may be millions of years ago) [Ray, 2008]; or, by analyzing a succession of sandstone and mudstone the geologists try to interpret if these rocks were deposited many million years ago from a river or were deposited below the sea [Leeder, 2011]. The sediments shed by the mountains like Himalaya is a storehouse of much information on the tectonic history of the mountain belt [Najman, 2006.]. Similarly, sedimentary rocks in Mars or moon provide important information on the history of these extra-terrestrial bodies [Beyer et al., 2012]. Some of the major spheres of research relevant to the study of sedimentary rocks are given below:

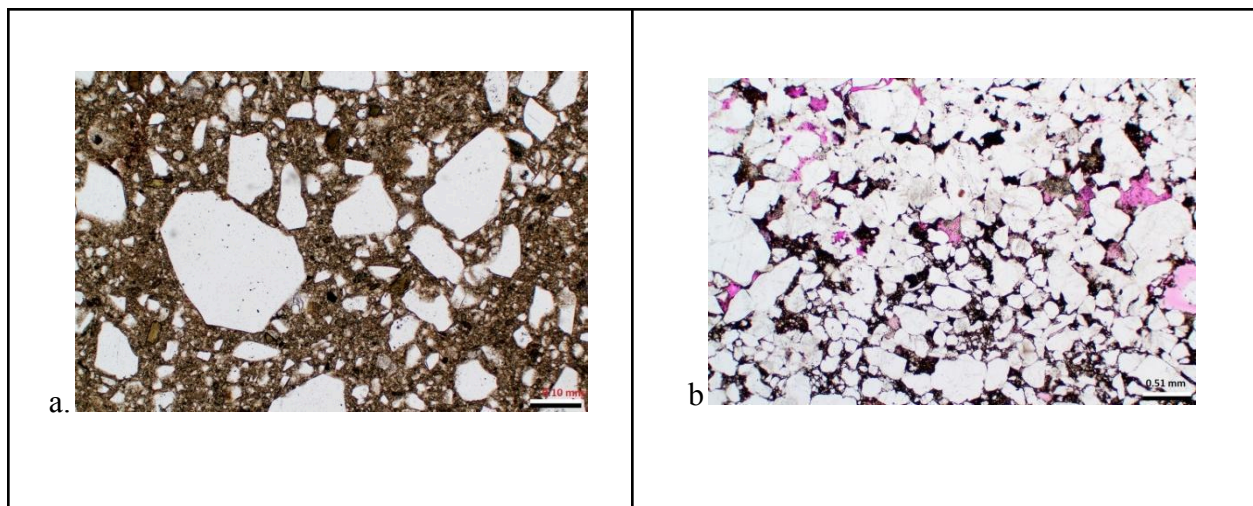
1. Explorations of Geo-Economic resources: Petroleum reservoirs have organic-rich, sedimentary source rocks that produce hydrocarbon under a specific temperature and pressure regime as these sediments are buried after deposition. Once the hydrocarbon is formed it tends to migrate to reservoirs of more porous rocks, mostly sandstone and limestone of different kinds. Similarly, groundwater is primarily stored in sandstone and is referred to as aquifer. Clastic sedimentary rocks also host many economic minerals such as gold and diamonds, which are often eroded from their source rocks and concentrated as placer deposits in the associated sedimentary rocks.
2. Environmental geology: Soils are produced due to the interaction of the rocks (sedimentary as well as other types of rocks) with the atmosphere. The soils that are formed in the geological past (few to many thousand or millions of years ago) are preserved in the rock record as ancient soils (referred to as paleosol). Paleosols are

sedimentary rocks and their identification and analysis provide clues to the ancient atmosphere and the climate [Tabor and Myers, 2015]. Thus paleosols, along with other proxies, are one of the important means of gaining insight into the ancient atmosphere. The ideas of ancient atmosphere and climate changes provide us a clue to their possible changes in future.

3. Significance in anthropological studies: In the archeological work concerned with human settlement, it is necessary to know about the sediments and their origin that occur with the buried ancient human settlements. Through the study of the comparatively younger sedimentary units it is possible to estimate the time when major paleogeographic/climatic changes took place in a particular location and how that influenced human habitation or establishment of human settlements or their abandonment like that in case of Harappa-Mohenjo-Daro civilizations [Singh et al., 2017], [Sarkar et al., 2020]. In order to understand the paleogeography and depositional environment, it is necessary to examine the texture and other parameters of sandstone and involves petrographic study of these sediments.
4. Our actions as humans have a strong effect on sedimentation and erosion. Construction of dams, embankments and other such activities produce adverse effects on the natural river systems. One way of monitoring the changes in the flow condition and sediment transport of such rivers is to examine the clastic sediments deposited by the embanked rivers. Understanding our impact on the environment and the environment's impact on us, must include deep appreciation for sediments and sediment transport [Ghosh & Chakraborty, 2022].

There are many methods and techniques used by the earth scientists for the analysis of clastic sediments and sedimentary rocks [Folk, 1980], [Tucker, 1988]. One of the important and crucial steps in this study of the detrital sedimentary rocks is to know the geometry and shape of the clasts (fragments of different minerals or rocks), also termed as grains, and the character of the fine-grained matrix (fine-grained, usually clay sized material or carbonates, that surround the individual sand-sized clasts) (Fig. 1.1 a, b). Geologists generally use a petrographic microscope to distinguish and study the clasts and the matrix. Traditionally the identified grains/clasts are

demonstrated by the geologist on a photomicrograph of the thin section captured on a polarizing microscope. Such images contain hundreds of grains/clasts of different sizes and characters. The manual delineation and identification of such grains is extremely time-consuming and tedious. Often the grain boundaries are fuzzy and their exact delineation is highly subjective at places. Therefore, the machine vision and image analysis techniques for grain segmentation from the matrix, and for the measurement of size and other properties of these grains, appear to be one of the potential tools. In the following, let us first explain very briefly the properties of the grains in sedimentary rocks.



**Fig. 1.1:** a. Matrix-rich sandstone b. Sandstone with little matrix in the interstices of framework grains.

### *1.2.2 Geological Significance of Grain Size Analysis in Sedimentary Rocks*

In geology grain size is defined as the average diameter of individual grains or particles of clastic sediments and rocks. Grain size is the primary parameter on the basis of which clastic sediments and rocks are divided into different classes. This subdivision is based on the Udden-Wentworth scale, which differentiates four major classes of sediments (and corresponding sedimentary rocks) that are further divided in sub-classes [Tucker, 2001]:

- a. Gravel or Conglomerate:



Grain size > 2 mm. It includes granules (2-4 mm); pebbles (2 – 64 mm), cobbles (64 – 256 mm) and boulders (> 256 mm). The corresponding rocks are conglomerate and breccia (further differentiated based on roundness).

b. Sand or Sandstone:

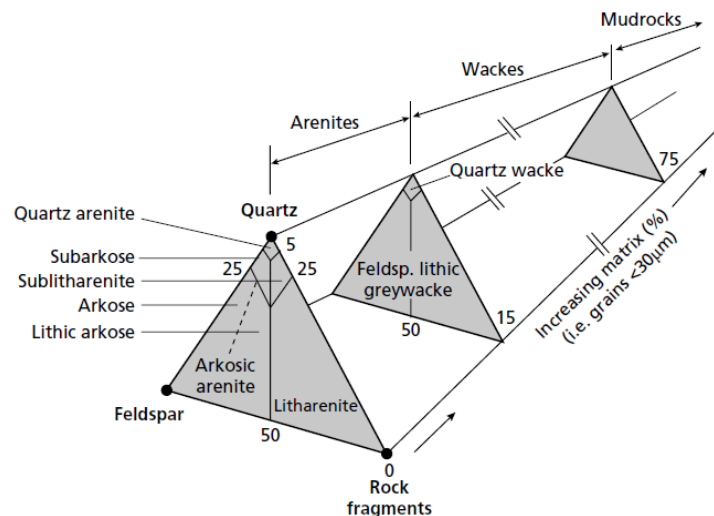
Grain size (0.0625 – 2 mm) and the corresponding sandstone.

(i) Silt or Siltstone:

Grain size (0.004 – 0.0625 mm) and the corresponding siltstone.

(ii) Clay or shale/clay-stone: Grain size (<0.0625 mm).

Based on the proportion of different minerals/rock fragments and the proportion of fine-grained matrix that constitute the clastic sedimentary rocks, they are further subdivided into several subgroups [Tucker, 2001] (Fig. 1.2).



**Fig. 1.2:** Classification of sandstone [Tucker, 2001]. Note the classification into wacke and arenite on the basis of proportion of matrix and framework grains.

There are two standard methods to define the grain size. As defined in the literature [Lopez, 2017], grain size (or clast size) is the longest diameter of individual non-spherical grains of sediment, or the lithified particles in clastic rocks. In theory, each particle constituting the sample is to be analyzed individually, and the final result is the combination of all the individual measurements. Particles are complex three-dimensional objects with specific lengths, widths,

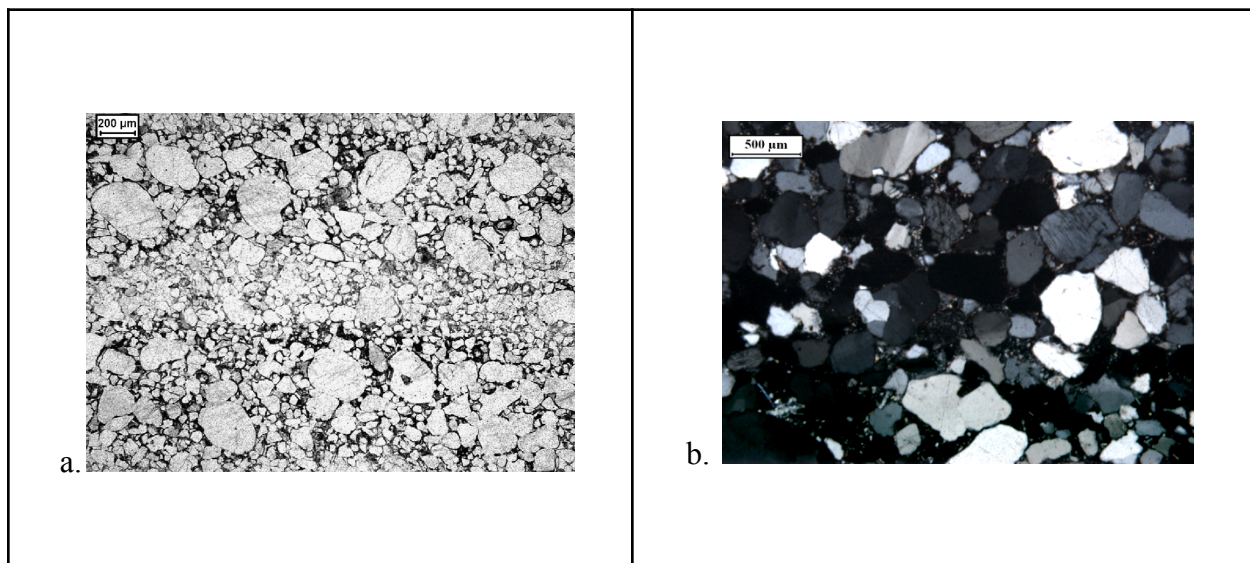
and thicknesses [Folk, 1980], [Stanley and Lines, 1992]. Only perfect spheres can be completely described by a single number, i.e., their radius or diameter. So, in order to simplify the measurement process, particles are most commonly conceptualized as one-dimensional spheres or circles rather than three-dimensional objects with irregular shapes. A cylinder with X length and Y diameter can be equivalent to a sphere of Z diameter that has the same volume as the cylinder. This is called the Concept of Equivalent Spheres [Jennings and Parslow, 1988]. This concept can be applied to a number of different grain size analysis measuring properties of a particle, such as maximum and minimum diameter, surface area, volume, or weight. Based on this concept of equivalent spheres, sedimentation rates can be calculated, and simple things such as sieve aperture sizes can be specified [Folk, 1980], [McManus, 1988]. The universally used scale for measuring the grain size is the Udden-Wentworth scale. It is a geometric scale of grain sizes which classifies particles of siliciclastic sediment from > 256 millimeters (boulders and may be meters in their longest dimension) in size down to 0.00006 millimeters (clay). This scale is almost universally accepted by geologists. It was first proposed in 1898 by Udden and then modified and extended in 1922 by Wentworth. Later on, in 1934 W.C. Krumbein [Krumbein, 1934] modified the Wentworth scale and proposed a new logarithmic scale computed using the equation,

$$\varphi = -\log_2 \frac{D}{D_0} \quad (1)$$

Where  $\varphi$  is the Krumbein phi scale,  $D$  is the diameter of the particle or grain in millimeters and  $D_0$  is a reference diameter, equal to 1 mm (to make the equation dimensionally consistent).

Identification or counting of minerals grains and measurement of their size in sediments or sands is an important and critical task in many scientific endeavors. To put a few referral examples from real life problems we can mention: (a) In order to understand the depositional environment and the regional paleogeography of a region in the past, it is very important to know the petrography of the sandstone and their texture. Aeolian sands are very well-sorted while that of the fluvial deposits are not so (Fig. 1.3a, b). (b) Type of the detrital grains in coarse clastic rock is important Hence, the study of geological thin sections (sedimentary petrography) has become one of the important to understand their provenance (the region from which the sediments were

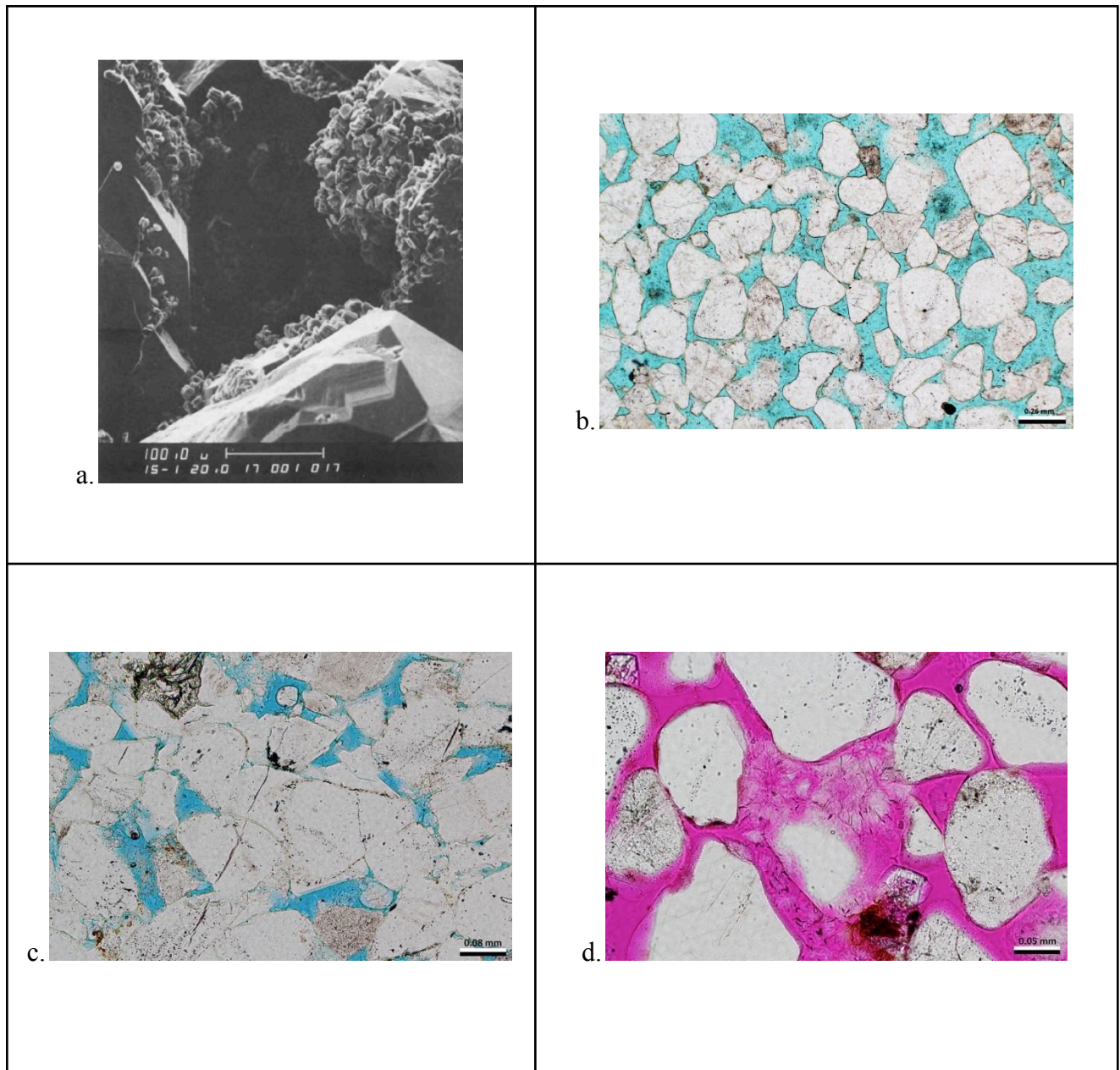
source), and the pathway of their transport. (c) Recovery of the oil and gas in the subsurface reservoirs are critically dependent on the texture, porosity or permeabilities of the rocks [Tucker, 2001]. In order to determine porosity-permeability it is necessary to identify the clasts and the matrix of the sandy rocks (Fig. 1.4a, b, c, d). There are several parameters that can define the rock texture: grain size, shape (sphericity), sorting, roundness (angularity), compaction (grain to grain contact) and preferred orientation of the grains [Folk, 1980]. Mean and median grain size, mode, sorting, and skewness are routinely determined by the geologists in order to understand the texture and provenance of the rock samples.



**Fig.1.3:** a. Thin section photomicrograph of a well sorted, well-rounded fine-grained Aeolian sediments. b. Photomicrograph of poorly sorted coarser sediments of fluvial deposit.

However, as already stated, visual identification of clasts of different minerals and the accurate estimation of their proportion is a lengthy, complex and error prone task that has to be performed by highly trained personnel. Only the sheer number of grains or particles to characterize a rock unit (typically in the order of several thousands to be statistically representative) render the operation tedious and time consuming. Therefore, use of computer vision and machine learning techniques to develop automated algorithms for the analysis of the images derived from microscopic studies of the clastic rock thin sections are becoming increasingly popular; but the first hindrance to this research and development is the absence of a realistic benchmark dataset

that the computer vision community may make use of to develop and further improve such algorithms which therefore constitutes the first important objective of this thesis.



**Fig. 1.4:** a: SEM photograph of (a clay mineral) growing on the quartz detrital grains in a sandstone, reducing its porosity; b. mature sandstone made of quartz and high porosity; blue color is an epoxy resin used to highlight the pore spaces; c. Reduction in the porosity of the sandstone due to development of silica cement; d. reduced porosity in similar sandstone with precipitation of the interstitial clay matrix.

### *1.2.3 Objective: Part I- A Novel Sandstone Photomicrograph Dataset Proposed as Benchmark*

There is currently a lack of a benchmark that includes a substantial number of segmentation datasets. The contemporary research papers utilize some proprietary datasets consisting of a limited number of three to four images, which are not accessible to the public. Additionally, they do not possess a realistic distribution of grain-matrix that has overlapping or fuzzy grain boundaries. A primary objective of this thesis is to address this gap in knowledge and develop a novel dataset, viz. SPGMS: Sandstone Photomicrographs for Grain-Matrix Segmentation which we propose as a valuable benchmark for segmenting the grain from the matrix.

The SPGMS dataset consists of nineteen sets of images, with each set containing plane polarized, cross-polarized, and corresponding ground truth images carefully labeled by the expert geologist. The benchmark was employed to assess the efficacy of multiple cutting-edge segmentation algorithms. The SPGMS dataset, which is a unique contribution of the present thesis, is publicly available at: <https://github.com/Rajdeep002/grain-matrix-dataset>. Since, there is some ambiguity with 2-3 images with respect to identification of some grains, for the present thesis, we finally considered sixteen images, though the remaining images may also be of use for future researchers.

### *1.2.4 Objective: Part II - Classical Computer Vision based Grain-matrix Segmentation*

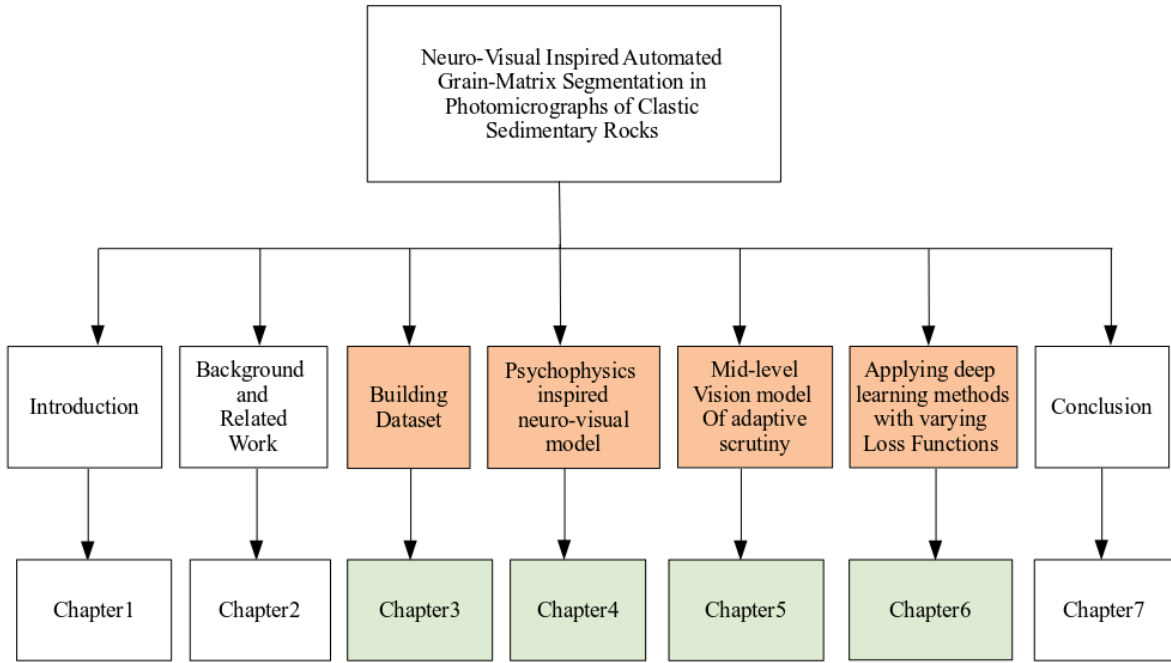
The present thesis aims to employ several classical machine vision-based techniques to analyze the photomicrographs of the clastic rocks taken under a petrological microscope and described above as our novel benchmark dataset contribution. These include several unsupervised methods like: Otsu [Otsu et al., 1975], Kapur's entropy [Kapur et al., 1985], Li's entropy [Li and Lee,

1993], Abutaleb's thresholding [Abutaleb and Eloteifi, 1988], Normalized Cut (NCut) [Shi and Malik, 1997] and Local Binary pattern (LBP) [Pietikäinen, 2010]. A significant contribution of the thesis is that it proposes some novel machine vision algorithms which are inspired by the neural processes in the central visual pathway of the mammalian/primate brain as is apparent from visual psychophysics based analysis of brightness-contrast perception occurring in illusory images according to human observers.

#### *1.2.5 Objective: Part III - Deep Neural Network based Grain-matrix Segmentation*

The objective of application of the very contemporary and state of the art deep learning techniques (also inspired by visual cortical processing) in computer vision, as yet another approach, is also practically the first of its kind as an application in this domain. We consider several deep learning algorithms as the supervised methods in this grain segmentation task that include: UNet [Ronneberger et al., 2015], UNet++ [Zhou et al., 2018], FPNNet [Esmailzahi et al., 2021], MANET [Li et al., 2021], LinkNet [Chaurasia and Culurciello, 2017] and PSPNet [Zhao et al., 2017]. Applying both the supervised and unsupervised methods (mentioned in section 1.2.4) enables the flexibility to analyze the dataset in greater detail.

Next, a chapter-wise enumeration of the background and contributions made in the thesis is presented in the following section, along with brief descriptions of each of the contributions. A schematic organization of the contributory chapters (Chapters 3-6) of the thesis can be found in Figure 1.5.



**Fig. 1.5:** Overall organization of the thesis

### 1.3 Thesis Organization and Chapter-wise Contributions

This thesis develops novel algorithms, both unsupervised and supervised: one in the framework of the parallel visual channels in human vision mechanism in an attempt to mimic the eye-brain system of the expert geologist behind the microscope; another one as a mid-level vision model that automates the visual pathway selection of the first algorithm. Besides, we have done extensive experiments with state of the art deep learning algorithms with a goal to arrive at a possible deep semantic grain segmentation including some novel experiments with the loss function. These algorithms will be detailed in the subsequent chapters, viz. Chapter 4 [Das et al., 2021], Chapter 5 [Das and Ghosh, 2022], and Chapter 6 ([Das et al. 2022], [Das et al. 2024]) respectively. Before that, in Chapter 2, we have introduced some of the related works along with the concepts of image analysis, visual perception, and neuro-visual mechanisms used for solving



the automated grain-matrix segmentation problem. While in Chapter 3, we have made a unique contribution in the form of building a proper dataset for this work, publicly available for researchers, the like of which was hitherto non-existent. A notable point to be mentioned here, is that the dataset building has been a continuous task during these works since appropriate slides for preparing the ground truth (explained in Chapter 3) has not always been readily available. The concluding remarks and future directions are presented in Chapter 7 followed by a list of publications from this thesis.



## **Chapter 2**

# **Grain-matrix Segmentation: Background and Related Works**

## **2.1 Introduction**

In the Earth and archaeological sciences, grain size analysis is a routine laboratory study. Particle or grain size is considered to be a fundamental attribute or physical property of particulate samples or sediments and sedimentary rocks. Much can be said from analyzing not only the size of clastic or detrital (inorganic), bioclastic (organic), or chemical particles but also from the overall size distribution, size fraction percentages, textural maturity of the sediment or sorting, surface texture attributes of a particle, and sphericity/angularity and shape of a particle. One of the particularities of grain size analysis is the importance of estimating correctly the different size fractions that constitute the ensemble of the material being analyzed. In theory, each particle constituting the sample is to be analyzed individually, and the final result is the combination of all the individual measurements. The most common process in grain analysis is the point counting method. Point counting involves looking at a large number of points on the thin section slide under the petrographic microscope (detailed in section 2.5), and recording exactly what is seen at each point by moving the slide by an equal increment along a line. Two important issues must be considered while studying rock texture using this conventional way under microscope: a) the experience and knowledge of the petrographer, and b) the time consumption. There lies the importance of machine intelligence based automated image analysis. Petrographic thin section image analysis involves segregating the mineral grains from the background, a task well-known as image segmentation in the machine vision domain. Several important aspects in relation to image segmentation and some related concepts in vision and perception are discussed in

subsequent sections of this chapter. First, we discuss the case of object-background segmentation, sometimes also referred to as figure-ground segmentation, which is especially relevant in the context of the present thesis. Then, we move on to expounding the other important concepts relevant to the present thesis.

## **2.2 Some Object-background Segmentation Techniques**

In many image analysis and computer vision tasks, image segmentation is a basic and crucial step. In some applications image segmentation is actually reduced to the form of binary classification and is referred to as figure-ground or object-background segmentation. Object-background segmentation plays a significant role in many important applications like document analysis [He and Schomaker, 2019], [QN et al., 2018], saliency detection [Hou and Zhang, 2007], video tracking [Chan, 2018] etc. The task of segmentation of the objects from the background is essentially to divide the image into two homogeneous partitions based on luminance, color, texture etc. This apparently simple task turns into an intriguing computational problem including the uncertainties and ambiguities with color, luminance and/or texture information when multiple objects are required to be detected from their background having overlapping boundaries. The problem becomes more acute when all the aforementioned situations are present in the image. One such situation arises in sandstone photomicrographs captured under a petrographic microscope.

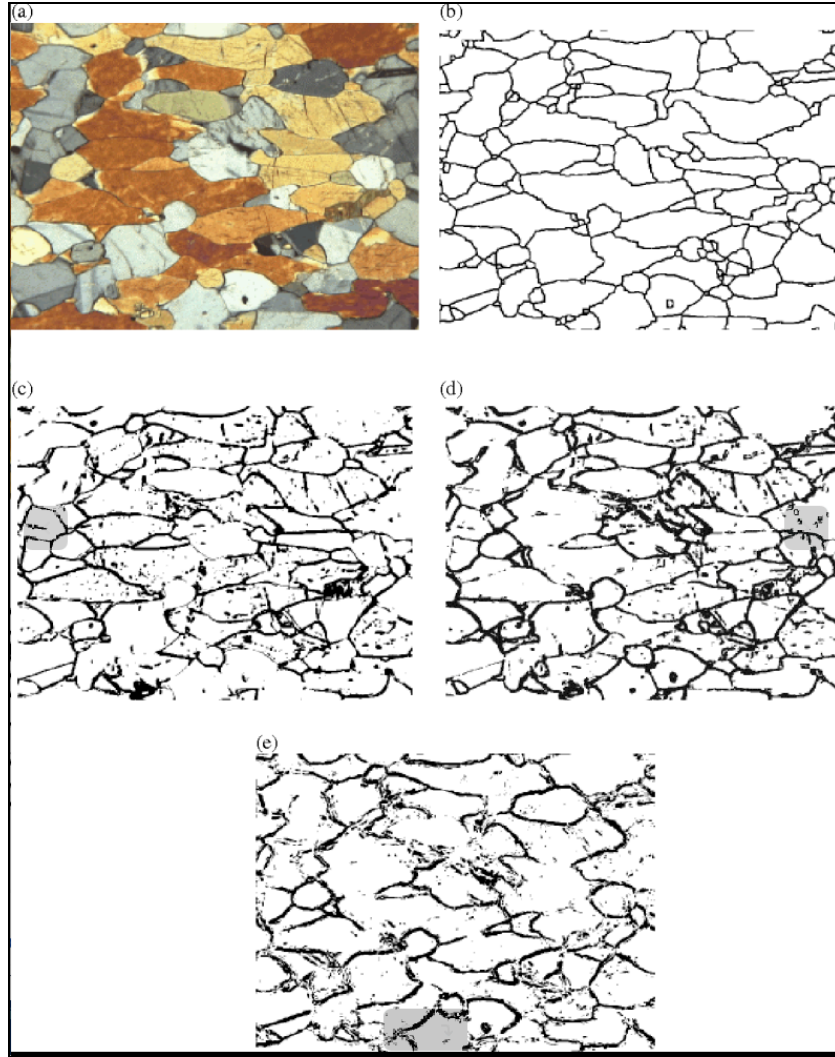
### ***2.2.1 Edge-Based Segmentation***

Edge-Based segmentation basically identifies the boundary of the objects by identifying the edges in the images using the gradient of image intensity. These can be a first or second order derivative operator, or an optimal edge detector. The gradient operators which involve first order derivatives respond whenever there is discontinuity in intensity level. These include the Prewitt, Sobel, and Roberts operators [Shah et al., 2020]. The second order derivative operators are particularly sensitive to the noise in an image, but can still be very useful, like for instance, the

Laplacian operator and Difference of Gaussian (DoG), and its equivalent Laplacian of Gaussian (LoG) are all rotation-invariant operators that detect edges in all directions by searching for zero-crossings [Eser and Derya, 2019]. The optimal edge detectors, such as Canny edge detector can produce continuous localized edges, and are devoid of multiple responses in presence of noise. It can detect strong as well as weak edges through two thresholds, the lower and the upper [Shah et al., 2020], [Canny, 1986]. In digital image processing with the clastic sedimentary photomicrographs, the edge-based segmentation [Goodchild & Fueten, 1998] targets to identify the grain outline using edge detection algorithms. An example of edge-based segment is shown in Figure 2.1

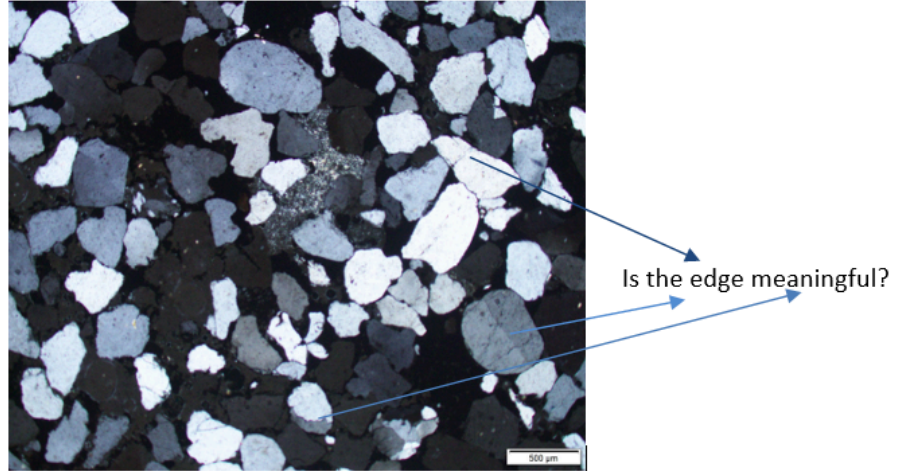
The edge-based segment, however, is weak due to several factors. The first factor is because using an edge operator, if all the edges present in an image are extracted then an additional decision making module is required to identify out of all the edges which corresponds to the boundary of the objects. It allows for missing and disjointed edges in separating the area of interest and background. For most of the cases, the edges that were produced by this technique were disjoint. The second factor is its noise sensitivity. As a result, for noisy images, the edges frequently failed to achieve correct segmentation. The third factor is its sensitivity to expanding of edges between important regions. These difficulties occur in images with high spatial resolution and complex geometric shapes. Consequently, the edges between regions are hardly defined, thus producing over or under segmentation. The sedimentary rock photomicrographs also represent a typical situation consisting of several uncertainties, and ambiguities.

These are not so apparent in the examples used in [Goodchild & Fueten, 1998]. If for instance, we look at Fig. 2.1, collected from that paper the segmentation performed seems quite robust and perfect; however, the image used here hardly contains any matrix region.



**Fig. 2.1:** Edge detection based grain segmentation [Goodchild & Fueten, 1998] using various edge detection techniques.

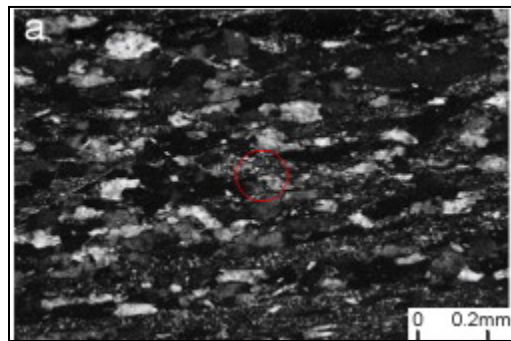
If we look at another image from our dataset (Fig. 2.2), the complexity of the problem becomes apparent. The three ambiguous edges will get detected through edge-detection algorithms, but none or some of them may not have any implication in grain-matrix segmentation.



**Fig. 2.2:** Edge detection method may not yield the desired segmentation as shown in the two ambiguous regions (This image is taken from our own dataset).

### 2.2.2 Energy-Based Segmentation

In energy-based segmentation methods, minimization of a pre-defined energy function is done. The methods are computationally costly besides being often unable to ensure convergence to global minima. Partial Differential Equation (PDE) based methods constitute an important plank in this category. PDE based methods have been widely and successfully used in image analysis. Using a level set [Osher and Sethian, 1988], for instance, [Chan and Vese, 2001] proposed a successful segmentation model (C–V model) that is based on the classical active contour model [Kass et al, 1988]. The C–V model basically allows to evolve an initial contour governed by a PDE till it stops on the edges of the target object. [Lu et al., 2009] employed the level set method to extract grain boundaries as shown in Figure 2.3.

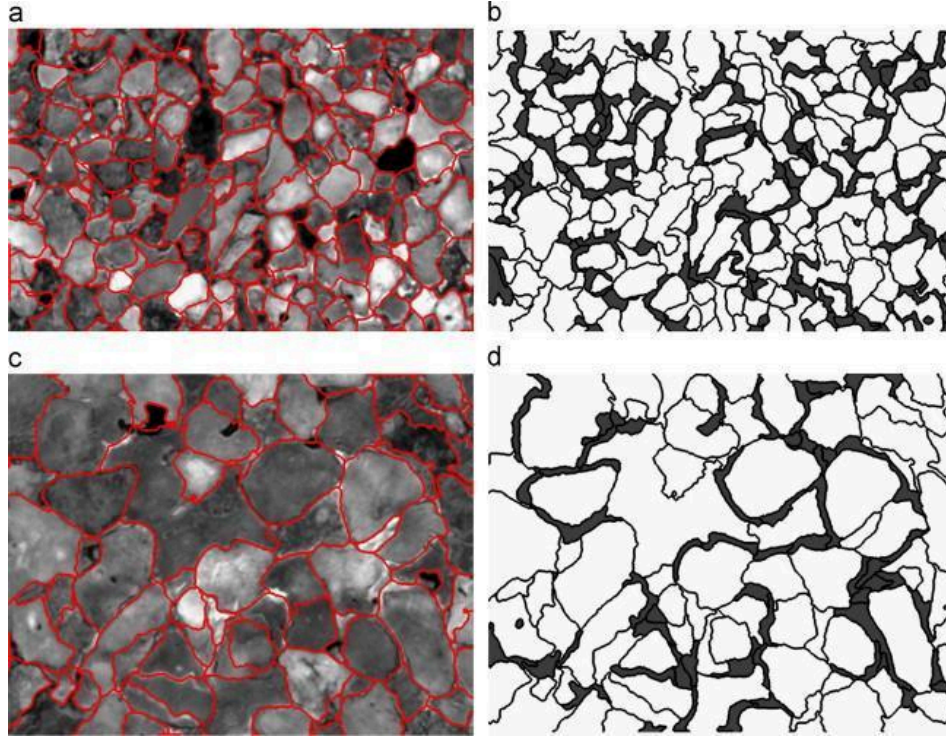


**Fig. 2.3:** Energy minimization based grain segmentation [Lu et al., 2009].

### *2.2.3 Region-Based Segmentation*

The region-based methods [Badrinarayanan et al., 2017], [Wei et al., 2020], [Wang et al., 2017], [Abutaleb, 1989] generally use clustering based approach utilizing the common properties of neighboring pixels, like gray value statistics, to isolate the grains. Some of the region-based methods also combine energy-based methods. For instance, [Zhu and Yuille, 1996] proposed an image segmentation technique based on a new region competition algorithm, which combines local pixel and region statistics for minimizing an energy function on an image scale. Here the adjacent regions compete for the ownership of common border points. Also, alongside competition, region merging is done through this approach. [Jungmann et al., 2014] successfully applied this region competition procedure on thin section images for identifying boundaries of quartz grains and anhydrite.

The region based methods also include the clustering technique. Clustering is applied in numerous fields of data mining, including geology. For images, the clusters may be created based on features like intensity, size, color and texture etc. K-means is a well-known algorithm which for image clustering is used for classifying image pixels into  $k$  numbers of clusters where  $k$  is a whole positive number. Very recently, [Izadi et al., 2020] proposed a novel incremental-dynamic clustering algorithm which is yet to be tested to prove its efficiency to correctly identify the grain pixels which are commonly identified as background pixels in geological photomicrographs. However, the clustering algorithms like k-means, for instance, also suffer from several drawbacks, such as lacking the right criteria to define the number  $k$  in this example. Results are also not the same when execution is repeated and high dependency on the initial conditions is to be noted. In some of the methodologies proposed in the present thesis we have used the k-means clustering technique in combination with visual perception based methods finally leading to figure-ground segmentation of grain and matrix in the photomicrographs.



**Fig. 2.4:** Segmentation has been done through region competition [Jungmann et al., 2014]. Classification has been performed into quartz (white) and anhydrite (black)

#### 2.2.4 Threshold-Based Segmentation

In threshold based segmentation approaches the difference in the color or grayscale values of the objects and the background is the main attribute used to divide the image into sub regions. The input to a thresholding operation is typically a grayscale or color image. In the simplest implementation, the output is a binary image representing the segmentation. The success of any thresholding based segmentation algorithm therefore strongly depends on the selection of threshold. The thresholding method has also been used in some of the proposed techniques of the present thesis. The Otsu (1979), Sahoo et al. [Sahoo et al, 1997], [Abutaleb, 1989], [Kittler and Illingworth, 1986], [Tulasigeri and Irulappan, 2016], Normalized Cut (NCut) [Shi and Malik, 1997] are some of the well-known threshold based segmentation algorithms. The results are not very satisfactory. This is because, in many real life applications, as also in the grain-matrix segmentation problem, distinguishability of pixels representing the object class and those representing the background class is often very poor [Wang et al., 2017]. Neither the distribution

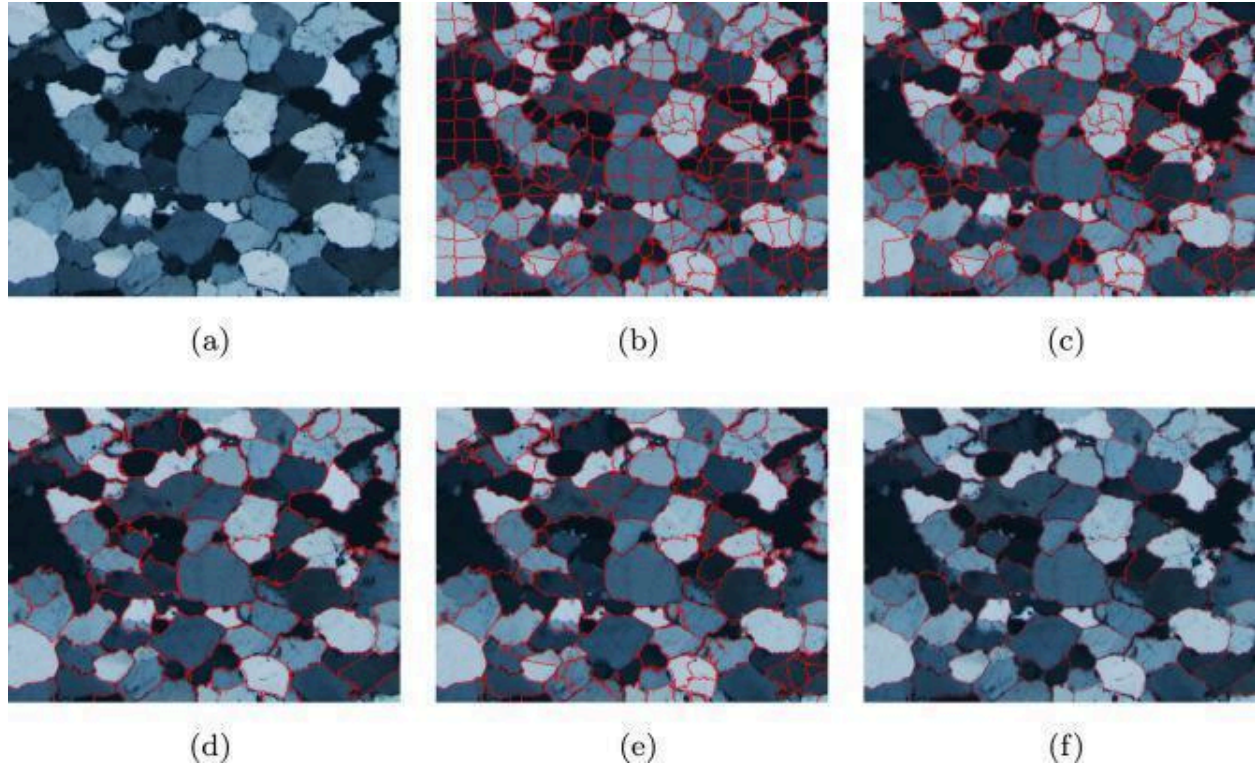
of grayscale or color pixel is fairly binary, nor is a distinct plateau observed in the distribution which can be considered as threshold. Selection of threshold for segmenting image into object and background still remains an active research topic.

### *2.2.5 Visual Perception-Based Segmentation*

David Marr's proposal of a computational paradigm for studying the visual system, wherein aspects of vision is described through a computational–reductionist approach brought a remarkable change in vision research. In Marr's proposed framework, at various stages of vision different representation schemes evoke and they are termed as low-, mid- and high-level vision [Marr, 1982]. Briefly the framework of Marr can be summarized as, Low-level vision, such as edge and motion detection, is explained by basic low-level neural circuits, whose outputs serve as building blocks for more complex circuits computing high-level features such as shape and entire objects. In low-level vision mostly intensity and chromatic features are analyzed. One of the important steps in the hierarchical processing through a combination of low-level and mid-level vision is the segmentation. Mid-level perception inspired superpixel based segmentation has already been utilized in image processing in general, and grain identification in particular. The goal of this method is to group pixels into perceptually meaningful regions, called superpixels. [Felzenszwalb and Huttenlocher, 2004], for instance, is a graph based superpixel algorithm, which maps the image to a graph and clusters the pixels by building a minimum spanning tree. [Van den Bergh et al., 2012] on the other hand, starts from a coarse partition and continuously refines the superpixels using hill-climbing method, SEEDS (Superpixels Extracted via Energy-Driven Samplings), to optimize an energy function defined on the similarity of color histogram. The SLIC (Simple Linear Iterative Clustering) [Achanta et al., 2012] algorithm uses k-means for localizing the pixel searches to limited regions around the centers with the distance defined on both color and spatial spaces. Later, [Jiang et al., 2018] proposed for petrographic microscopy what he termed a multi-angled SLIC (MSLIC). In the second stage of this algorithm, visual perception was further explored and a two-step merging



(CFM) technique was used to merge the superpixels according to the Gestalt laws [Desolneux et al., 2007], which contains a coarse merging (CM) algorithm and a fine merging (FM) algorithm, much like what happens in the two main parallel visual channels, the Magno and the Parvo. This concept, we have also utilized in the present thesis, to develop two algorithms in the framework of human vision mechanism at mid-level, the specific concepts being elucidated in a later section. Here some results from [Jiang et al., 2018] are reproduced.

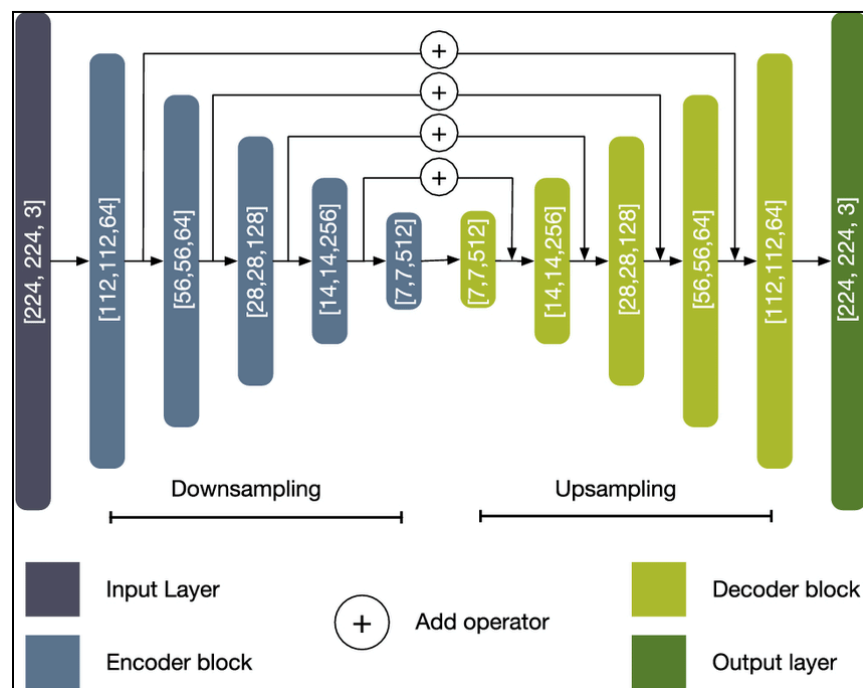


**Fig. 2.5:** Visual comparison among different methods on quartz sandstones. (a) the original microscopic image. (b) the segmentation result of MSLIC. (c) the segmentation result of CM. (d) the segmentation result of CFM. (e) the segmentation result of MRM. (f) the manually segmented ground truth [Jiang et al., 2018]. CFM consists of CM and FM and is compared with MRM, an algo developed earlier by the same authors.

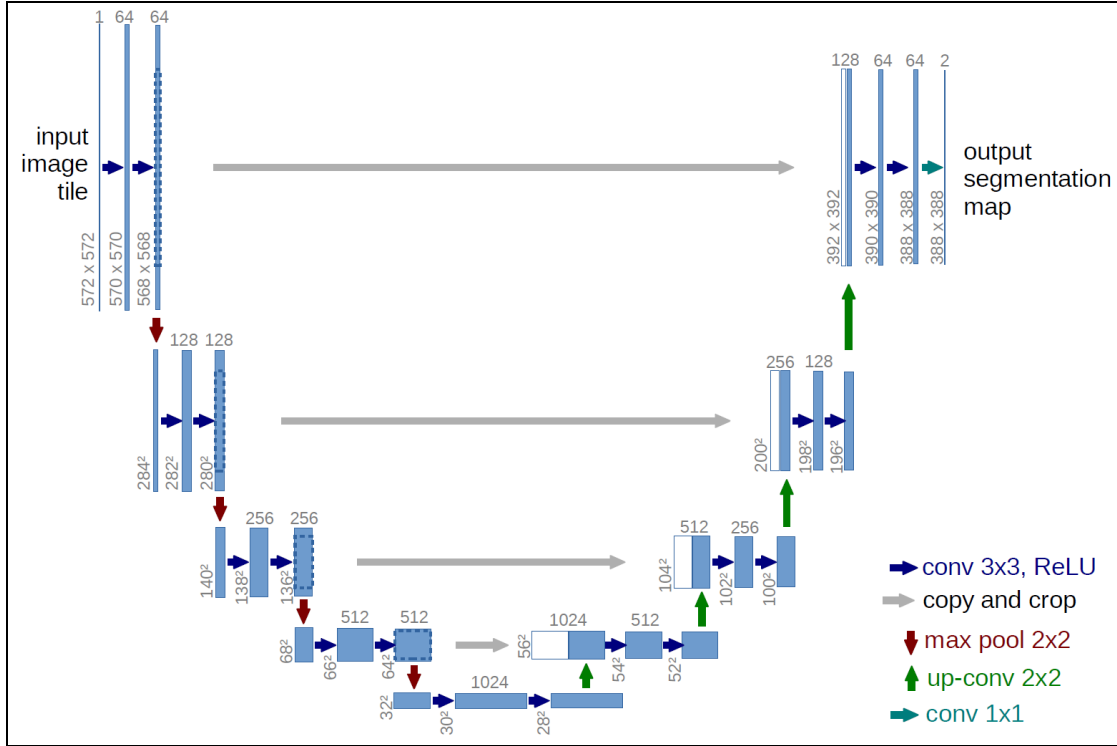
### 2.2.6 Deep Learning-Based Segmentation

To the best of our knowledge, deep learning based approach to grain-matrix segmentation has not been widely applied. Only very recently, almost simultaneously when our work in this respect was published online in 2021 [Das et al., 2022], that a work on CNN based semantic segmentation was published [Saxena et al., 2021] which investigated the feasibility of employing the CNN in a slightly different problem of obtaining binary pore as well as a ten-class semantic

segmentation. One reason for lack of more literature in deep learning may be the dearth of dataset, but the fact remains that the present thesis can possibly be noted as the first such attempt of binary segmentation of clastic sedimentary rocks through deep neural networks. Of course, there are several instances in computer vision where deep learning has already been effectively employed to carry out image segmentation. A comprehensive review was very recently published in this domain [Minaee et al., 2022]. Some of the deep architectures commonly used in image segmentation and relevant to our contribution in this thesis (Chapter 6) are shown below. Details will be discussed in the relevant chapter later. Finally, an important research on sedimentary rock analysis based on Deep Neural Nets was published in the Journal of Sedimentary Research (2023) by Murtazin et al.



**Fig. 2.6:** The Linknet [Garcia et al., 2020] architecture



**Fig. 2.7:** The U-net [Ronneberger, 2015] architecture.

## 2.3 Some image segmentation techniques: Supervised, unsupervised and self-supervised

As already explained, image segmentation is the process of dividing a picture into distinct segments in order to identify and distinguish objects, boundaries, and regions within the image. This method is essential in a wide range of applications, including medical imaging [Patil et al., 2013], autonomous driving [Kaymak et al, 2019], object identification [Barik et al, 2010], and other fields [Kaur et al., 2014, He et al., 2024]. Based on whether a-prior labeled data are used to train the segmentation model, segmentation methods on images can be broadly categorized into three primary types - supervised, unsupervised, and self-supervised methods of segmentation. For example, the edge-based, region-based, and energy-based segmentation methods, described previously, as also the novel visual perception based methods proposed in this thesis, fall into the category of unsupervised methods. Deep learning-based segmentation, including those proposed

in the thesis, belongs to the supervised category. Below we give a brief description of these three types.

### *2.3.1 Supervised image segmentation techniques*

Supervised image segmentation methods entail training models using annotated data to categorize each pixel into specified classes. Deep learning methodologies, particularly Convolutional Neural Networks (CNNs), encompassing Fully Convolutional Networks (FCNs) and U-Net, are extensively employed for their capacity to learn intricate spatial hierarchies and generate pixel-level predictions [Long et al., 2015; Ronneberger et al., 2015]. Support Vector Machines (SVMs) categorize pixels utilizing feature vectors obtained from pixel intensities and textures [Schölkopf et al., 2001], whereas Random Forests (RF) amalgamate several decision trees to enhance segmentation precision via majority voting [Breiman, 2001]. K-Nearest Neighbors (K-NN) classifies each pixel by evaluating its feature vector against those of its nearest neighbors in the feature space [Cover & Hart, 1967]. Conditional Random Fields (CRFs) are frequently utilized as a post-processing technique to enhance segmentation outcomes by modeling pixel interdependencies and promoting smooth borders [Krähenbühl & Koltun, 2011]. Some recent supervised image segmentation works can be found at [Pont et al., 2015, Han et al., 2024, Aljuaid et al., 2022]. As mentioned previously in this chapter, some results, applying this type of method, have been reported in chapter 6 of this thesis.

### *2.3.2 Unsupervised image segmentation techniques*

Unsupervised image segmentation techniques aim to partition an image into meaningful regions without the use of labeled data. Common methods include K-Means clustering, which groups pixels based on feature vectors like intensity, and Mean Shift clustering, a non-parametric approach that identifies regions of high pixel density without predefining the number of clusters [Lloyd, 1982; Comaniciu & Meer, 2002]. Graph-based methods like Normalized Cuts and Graph Cuts segment an image by optimizing a cost function that balances within-segment similarity and between-segment dissimilarity [Shi & Malik, 2000; Boykov & Jolly, 2001]. The Watershed algorithm uses topographic concepts to identify regions by simulating flooding from local minima, although it may over-segment [Meyer, 1992]. In a very recent work on sedimentary rock segmentation problem, Murtazin et al. (2023) have applied this unsupervised technique and

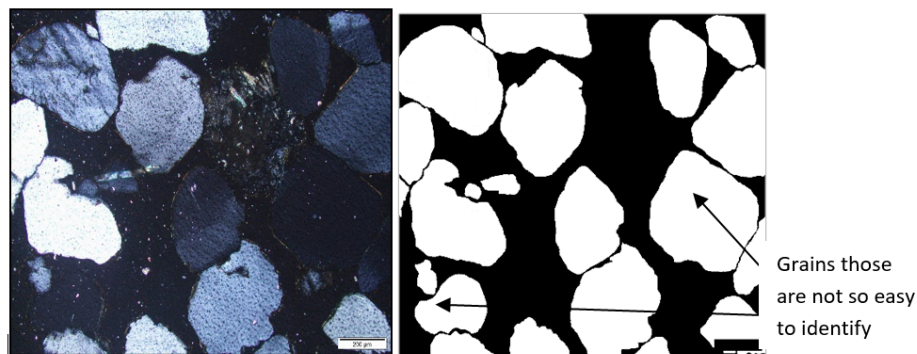
compared the same with some supervised methods. Self-Organizing Maps (SOMs) reduce dimensionality and group similar pixels together for segmentation [Kohonen, 1995], while texture-based methods, such as Local Binary Patterns (LBP) and Gray-Level Co-occurrence Matrices (GLCM), classify pixels based on texture features to separate regions with distinct textures [Ojala et al., 1996; Haralick et al., 1973]. These unsupervised techniques are widely used in applications like medical imaging, remote sensing, and object recognition. Some results applying this type of method on our novel dataset have been reported, in the later portion of chapter 3 of this thesis.

### *2.3.3 Self-supervised image segmentation techniques*

Self-supervised image segmentation techniques aim to learn meaningful image representations without requiring labeled data by leveraging pretext tasks that generate supervisory signals from the data itself. These methods often involve training models to solve auxiliary tasks, such as predicting pixel or patch-level information, which forces the model to understand underlying structures and spatial relationships in the image. One prominent approach is contrastive learning, where models learn to distinguish between similar and dissimilar image patches by maximizing the similarity between positive pairs (augmented versions of the same patch) and minimizing the similarity between negative pairs (different patches) [Chen et al., 2020]. Deep clustering methods, such as DeepCluster, cluster image features and assign pseudo-labels to train the network, iteratively refining the segmentation model [Caron et al., 2018]. Generative models, like autoencoders and generative adversarial networks (GANs), have also been used to learn representations by reconstructing the input image or generating new ones, helping to capture fine-grained details necessary for segmentation [Kingma & Welling, 2013; Goodfellow et al., 2014]. In some approaches, self-supervised pre-training on large unlabeled datasets followed by fine-tuning on small labeled sets has been effective in achieving high segmentation performance [Xie et al., 2021]. These techniques have gained popularity as they reduce the reliance on expensive labeled data while still achieving competitive performance in segmentation tasks, particularly in domains like medical imaging, satellite imaging, and autonomous driving. Although this thesis does not elaborately study the application of self-supervised techniques on sedimentary rocks, we have, as an instance, in Appendix C, shown the results from one of these methods [Lin et al., 2023], when applied to a sample input cross-polar image from our dataset.

## 2.4 Some concepts in Neuro-visual and psychophysics algorithms used in this thesis

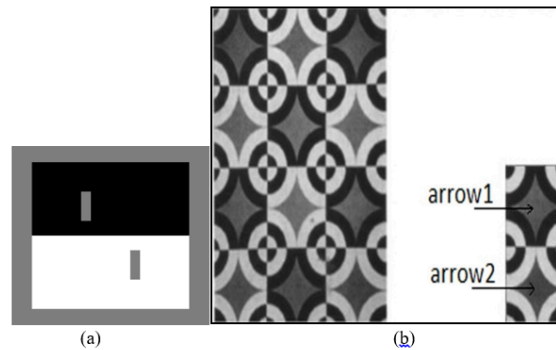
The brightness of any surface depends on two factors, the strength of the source by which the surface is illuminated and the reflectance of the material of the surface. Here we cite an example of a cross-polarized image from our dataset to illustrate how brightness perception may affect grain identification and counting.



**Fig. 2.8:** A typical sedimentary rock photomicrograph and its ground truth marked by the expert. From a comparison between the two, it is clearly evident that there is a wide variation of brightness of the grains, and a couple of these, marked by arrows, are so dark that they are practically invisible in the cross-polar images.

Actually, the perceived brightness of any surface depends on the brightness of the surface that surrounds it, and this phenomenon is termed as brightness induction [Kingdom, 2011]. Studies of [Helson, 1963] and also [Jameson & Hurvich, 1989] reveal that brightness-related perception is a complex task in the brain [Kingdom, 2011] and is realized as a result of brightness induction occurring from the neighborhood of any region of interest (ROI). The reason behind such complexity is that such induction is no way unidirectional. Further, brightness contrast and brightness assimilation are two opposite bidirectional aspects of brightness induction that arise in our vision. The brightness contrast is believed to appear by filtering out the low spatial frequency components of the stimulus while the brightness assimilation arises due to filtering out of high frequency components [Blakeslee and McCourt, 2004]. In the following we represent the

examples of brightness contrast and assimilation in Figure 2.9. In Figure 2.9(a), it is shown that the two test patches of equal intensity are perceived to be of different brightness due to the presence of surroundings, such that the patch on a darker background appears brighter than the other one on a lighter background. In Figure 2.9 (b) the brightness induction appears to be in the opposite direction. In recent past several attempts were made to model the complex psychophysics of brightness computationally [Blakeslee and McCourt, 2004], [Ghosh, 2012], but it is rare to find instances that the models are applied successfully to perform complex brightness perception-based image analysis tasks like geological photomicrographs or in medical images. Further, the corresponding models that are analogous to the neural networks active in the central visual pathway [Ghosh, 2016], also find few applications, except for a few general ones related to simple tasks like edge/line detection [Marr, 1982] or more complex tasks like object semantic representation [Wei et al., 2020], [Wei, 2016]. These have rarely been exploited to perform critical classification tasks like object segmentation although the same is done visually by experts in medical science or geological sciences using varied imaging modalities.

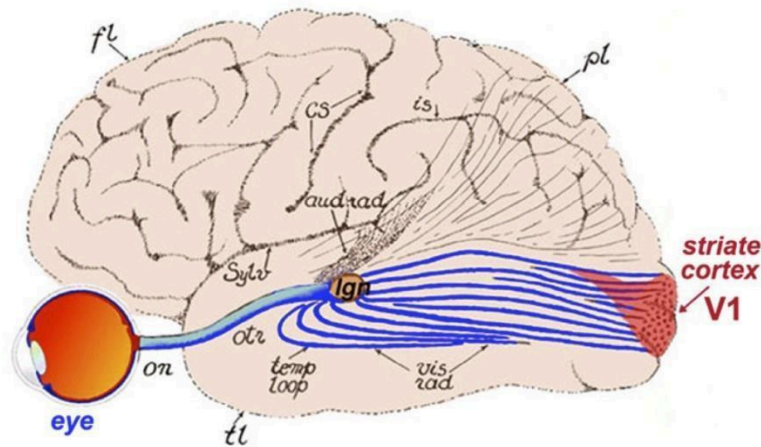


**Fig. 2.9:** Examples of (a) brightness contrast and (b) brightness assimilation. In (a) the two test patches of equal intensity appear to be of different brightness based on the surroundings in such a way that one on a darker background appears to be of brighter shade than the other one on a lighter background. In (b) the brightness induction is opposite. The surrounding darkness or lightness seems to have been assimilated into the centremost patches (marked by Arrow 1 & Arrow 2) thus representing a phenomenon directionally opposite to contrast. [De Weert & Spillman, 1995]

One of the central quests searched in the present thesis is to explore the parallel channels, especially how the Magnocellular (M) and the Pervocellular (P) channels in the central visual pathway (Fig. 2.10) can play a significant role in performing automated image segmentation of



the cross polar photomicrographs, by designing three filters. In the proposed computational model we tried to show that an expert's eye-brain system segments the grain under a petrographic microscope, via vision at a glance (VG), vision in strong contrast mode (VSC), and vision with scrutiny (VS), the third one being a combination of the first two, slightly modifying the proposal of [Hochstein and Ahissar, 2002]. In Chapter 4 of this thesis, initially to substantiate our claim we examined a few cases of optical illusion related to brightness perception using a proposed neuro-visual model. Thereafter, the investigation progressed to develop psychophysics inspired automated figure-ground grouping tasks for the grain and the matrix in cross-polarized photomicrographs. The algorithms and the computational details will be presented in the relevant chapter.



**Fig. 2.10:** The parallel visual channels from eye to the striate cortex/primary visual cortex (V1). The two principal divisions in these parallel channels are the Magnocellular (M) with lower spatial resolution, named Vision at a Glance (VG) and the Parvocellular (P) with higher spatial resolution, named Vision in Strong Contrast mode (VSC) in this thesis. [Image courtesy: <https://webvision.med.utah.edu/> ]

## 2.5 Related Works in Computer Vision based Petrographic Microscopy

It has already been mentioned at the beginning of this chapter that in geology (and also in archaeology), grain size analysis is a routine laboratory study. Particle or grain size is considered



to be a fundamental physical property of particulate samples or sediments and sedimentary rocks [Folk, 1980], [Friedman and Sanders, 1978]. The size of clastic or detrital particles (that include mineral or rock fragments and also fragments of animal shells), the overall size distribution, size fraction percentages, manner of packing, textural or mineralogical maturity, and also the shape of the particles and their surface features [Krumbein and Sloss, 1963], [Syvitski, 2002] are studied because all these features provide important clue to the genesis of the particles and the sediments. One of the particularities of grain size analysis is the importance of estimating correctly the different size fractions that constitute the ensemble of the material being analyzed. In theory, each particle constituting the sample is to be analyzed individually, and the final result is the combination of all the individual measurements. The most common process in grain analysis is the point counting method. It is a technique used in petrography to statistically measure and establish the proportion of constituents and minerals in coarse grained igneous, metamorphic, and sedimentary rock samples. In case of clastic sedimentary rocks, point counting is normally conducted manually and visually [Delesse, 1848], [Shand, 1916], [Chayes, 1949]. In a polarizing microscope, a stepper stage and a recording device is attached. In this technique the study slide is divided into grids, so that one can have 300-1000 grids. The size of the grids or the spacing between grid intersection points are chosen depending on the size of the grains in the particular sample being studied. At each grid intersection position, the petrographer notes the mineralogy of the grains and may measure shape or size of the detrital particles. In other words, point counting involves looking at a large number of points on the thin section slide, and recording exactly what is seen at each point by moving the slide by an equal increment along a line. At the end of each line the petrographer has to reset the stepper stage again at the start of another line, recording each new grid point. Two important issues must be considered while studying rock texture using the conventional way under a microscope: the experience of the petrographer and the time required for counting a minimum of 300 grains from each thin section. Although the method offers quite reliable results, it is time consuming and requires extensive geologist/petrographer effort. In addition, point counting results are subjective and depend on the petrographer's knowledge and expertise. Therefore, it is highly desirable to develop techniques and tools that provide reliable estimation of grain texture, while simultaneously decreasing the operating time and increasing the analytical accuracy. Although there are automated image processing techniques for counting features on microscope slides in other fields, particularly in

biology, these techniques cannot be easily adjusted to the petrographic studies of sediment or sedimentary rock thin sections. Because of the characteristics of petrographic slides in geology, the color of each grain is not uniformly homogeneous to be interpreted as a single component by an automated image processing technique, based on color or gray scale images. Precisely, images obtained from rock thin sections show inhomogeneous colors for each mineral type due to subtle tone changes related to inhomogeneous chemical composition of all natural minerals. Also, the color or gray-scale shade can be different between the grains of the same mineral due to deformation or stress to which the rock has been subjected. Besides, in all the natural rock thin sections, the crystallographic orientation of the minerals differs, and as a result the color of the same adjacent minerals can be different. Also, the birefringence (the color of a mineral when observed through cross-polarised light in a petrographic microscope) differs due to their varying orientation in thin sections. The birefringence color of the different minerals of the same species can differ also due to stress or diagenetic alteration affecting the rock. Obtaining grain texture from petrographic images has, for the reasons outlined above, long been an attractive but difficult challenge [Maitre et al., 2019]. Besides petrographic image analysis there are other analytical techniques namely, sieving analysis, Analysis using sedimentation or settling, laser diffraction method, dynamic light scattering and many others. A good review on the topic is available in [Lopez, 2017]. Recently a number of computer- aided segmentation methods have been published [Maitre et al., 2019], [Budennyy et al., 2017] and they tried to solve the problems associated with petrographic microscopy-based image analysis. For example, Machine learning and computer vision based tools proposed by [Maitre et al., 2019] shows promising results. Overall, the machine aided methods of grain segmentation may be categorized broadly into four main classes namely, (1) edge-based, (2) energy minimization based, (3) region or clustering based and finally (4) models based on visual perception. The edge-based segmentation [Heilbronner, 2000] is the traditional approach of image processing in which the target is to identify the grain outlines using edge detection algorithms. In energy-based segmentation methods [Lu et al., 2009], [Jungmann, 2014] minimization of a pre-defined energy function is made. The methods are computationally costly besides unable to ensure convergence to global minima. The third methods i.e. the region-based methods [Yesiloglu et al., 2012] [Izadi et al., 2015] [Izadi et al., 2017] [Izadi et al., 2020] generally use clustering based approach utilizing the common properties of neighboring pixels to isolate the grains. Very recently Izadi et al., [Izadi et

al., 2020] proposed a novel incremental-dynamic clustering algorithm which is yet to be tested to prove its efficiency to correctly identify the grain pixels which are commonly identified as background pixels in geological photomicrographs. Murtazin et al. (2023), in another recent work, used a high precision algorithm for grain segmentation of thin sections by multi-angle optical-microscopic images.

## 2.6 Some metrics used for performance evaluation in image segmentation

Image segmentation performance is evaluated using some evaluation indices, particularly e.g. Dice score, Accuracy, Jaccard index, Precision, Recall etc.

The Sorensen-Dice coefficient, it is well-known, is a basic and simple statistic used for comparing the similarity between ground truth and the segmented image. It is defined as:

$$\text{Dice score} = \frac{2 |X \cap Y|}{|X| + |Y|} \quad (3)$$

Another well-known validation index to compare segmentation results, viz. Accuracy, is defined as:

$$\text{Accuracy} = \frac{|X \cap Y| + |\bar{X} \cap \bar{Y}|}{|X \cap Y| + |\bar{X} \cap \bar{Y}| + |X \cap \bar{Y}| + |\bar{X} \cap Y|} \quad (4)$$

The third validation index, also frequently used for comparing segmentation results is the Jaccard index that we consider here as a third metric for analyzing the present grain segmentation results. It is defined as:

$$\text{Jaccard index} = \frac{|X \cap Y|}{|X| + |Y| - |X \cap Y|} \quad (5)$$

Precision and recall are defined as

$$\text{precision} = \frac{|X \cap Y|}{|X \cap Y| + |\bar{X} \cap \bar{Y}|} \quad (2)$$

$$\text{recall} = \frac{|X \cap Y|}{|X \cap Y| + |X \cap \bar{Y}|} \quad (3)$$

, where  $X$ ,  $Y$ ,  $\bar{X}$ ,  $\bar{Y}$  are the white pixels in the ground truth, the white pixels in the segmented image, the black pixels in the ground truth, the black pixels in the segmented image respectively.

## 2.7 Conclusion

Significant advances have been made in the domain of computational technologies on one hand and image processing and analysis techniques, on the other, especially in the last three decades, and consequently, several efforts have been made to deal with the problem of automatic grain segmentation of rock thin section images. The task of grain-matrix segmentation has been approached effectively through two alternative paths: grain boundary detection and direct segmentation of the grains from their background, i.e., the matrix zone in the rock thin section. However, the main objective for the approaches is the same, viz. the identification of grains. Most of these works, described above, have been performed with a very small number of images (often two or three), and despite all these efforts, there exists, to the best of our knowledge, no publicly available dataset of clastic sedimentary rocks for grain-matrix segmentation against any of the published papers. Most of the codes referred to in the papers, which are a few years old, are also currently not available in the links provided in the paper. Hence, during the present thesis, we decided to build a dataset for this purpose, and make it publicly available. In the next chapter, we shall be describing the dataset that we have built, and the process of preparing it, which is one of the contributions of this thesis.

# Chapter 3

## Building a Dataset of Clastic Sedimentary Rocks for Segmentation: a Unique and Timely Contribution

### ***Publication:***

*Das, R., Chowdhury, A., Chakraborty, T., & Ghosh, K. (accepted for publication in SPIE proceedings). SPGMS: An Annotated Benchmark for Grain-Matrix Segmentation in Sandstone Photomicrographs. ICVIP 2024 (The 8th International Conference on Video and Image Processing) which will be held on December 13-15, 2024 in Kuala Lumpur, Malaysia.*

### **3.1 Introduction**

In the previous Chapter (i.e. chapter 2) of this thesis, we have demonstrated several examples of Grain-matrix segmentation from literature (e.g. Fig. 2.1, Fig. 2.3, Fig. 2.4, Fig. 2.5). It may so seem that such segmentation has been done in those works based on some publicly available datasets. However, one may be disappointed to know that till date there is no such publicly available dataset of photomicrographs of clastic sedimentary rocks at present which can be used by the Computer Vision and Machine Learning community for developing and testing any new algorithm for grain-matrix segmentation. All the published papers in this domain that we found used their own respective datasets (which we also found to be mostly very small i.e. hardly about 3-4 images or even less) with corresponding ground-truth, but did not make those images publicly available. In Table 3.1, we have highlighted some of these relevant papers. It was with this serious handicap, or one may say a big gap in research in absence of any benchmark dataset that the present research work was initiated. So the first challenge was to build a dataset of a reasonable number of photomicrographs to bridge this gap. From getting a good slide from the

geologists to developing the corresponding ground truth, this is a tedious process, briefly described in this chapter, that continued throughout the work during this thesis.

| <b>Paper</b>                         | <b>Work in brief</b>                                                                                                  | <b>No. of images used</b> | <b>Dataset availability</b> | <b>Remarks</b>                                                 |
|--------------------------------------|-----------------------------------------------------------------------------------------------------------------------|---------------------------|-----------------------------|----------------------------------------------------------------|
| [Lumbreras et al., 1996]             | Divided marble images acquired with a petrographic microscope into distinct segments by using delineation of regions. | 6                         | NPA                         |                                                                |
| [Vanden Berg et al., 2002]           | Autonomously segmented interconnected particle regions in digital images of granular material.                        | 10                        | NPA                         | Contains approximately 1000 grain sections.                    |
| [Zhou, Y., Starkey, J. et al., 2004] | Integrate edge detection and segmentation methods for the segmentation of petrographic images.                        | 600                       | NPA                         | Used 3 sets of thin sections and each set contains 200 images. |
| [Mingireano v Filho et al., 2013]    | Used optimum path forest operators process for the task of sandstone segmentation.                                    | 151                       | NPA                         |                                                                |
| [Izadi et al., 2015]                 | Segmented minerals image based on incremental color clustering methods.                                               | 165                       | NPA                         |                                                                |

|                       |                                                                                                                                                                                |               |                                                                                                                                                                                                                      |                            |
|-----------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------|
|                       |                                                                                                                                                                                |               |                                                                                                                                                                                                                      |                            |
| [Izadi et al., 2017]  | Built an intelligent system for the sake of mineral identification in thin sections based image segmentation using an incremental clustering process and identification phase. | 135           | NPA                                                                                                                                                                                                                  |                            |
| [Izadi et al., 2020]  | Used an incremental-dynamic clustering approach for mineral segmentation.                                                                                                      | 16            | Alborz Mineralogical Database (Link: <a href="https://users.encs.concordia.ca/j_sadri/Alborz_Mineralogical_Database_V.1.0.zip">https://users.encs.concordia.ca/j_sadri/Alborz_Mineralogical_Database_V.1.0.zip</a> ) | 404 - page not found error |
| [Saxena et al., 2021] | Created a 10-class classification framework applying CNNs for thin section sandstone segmentation.                                                                             | 1300 +        | NPA                                                                                                                                                                                                                  |                            |
| [Das et al., 2021]    | Utilized psychophysics inspired models for sandstone thin section segmentation.                                                                                                | Not mentioned | NPA                                                                                                                                                                                                                  |                            |
| [Das et al., 2022]    | Developed a deep learning-based end-to-end pixelwise trainable framework named Deep Semantic Grain Segmentation network for the sake of sandstone thin section segmentation.   | 10            | NPA                                                                                                                                                                                                                  |                            |

|                                 |                                                                                                                             |      |     |                                                                                                                 |
|---------------------------------|-----------------------------------------------------------------------------------------------------------------------------|------|-----|-----------------------------------------------------------------------------------------------------------------|
| [Murtazin et al., 2023]         | Used high-precision algorithm for grain segmentation of thin sections by multi-angle optical-microscopic images.            | 18   | NPA |                                                                                                                 |
| [Chen, Y., Yi, Y. et al., 2024] | Combined petrology and image processing (concavity and corner detection) approach to segment sandstone thin section images. | 1000 | NPA | These images accommodates 653 groups of conventional sandstone images and 473 groups of tight sandstone images. |
| [Wang, C., Li, P. et al., 2024] | Introduced a machine learning viewpoint Rock-ViT in identifying complex lithological patterns.                              | 2257 | NPA |                                                                                                                 |

**Table 3.1:** List of related type data works. NPA: Not Publicly Available.

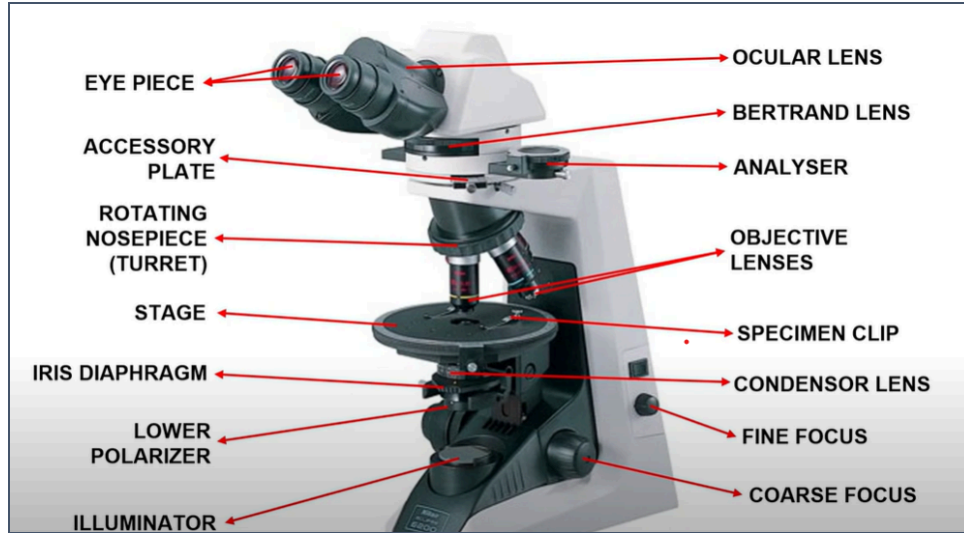
We have now built a dataset of 19 images, containing a cross polar (XPL) image, a plane polar (PPL) image, and the ground truth (GT) image. These different modes, viz. PPL and XPL are explained in the subsequent sections of this chapter. Since there is some ambiguity in a few of these, we have finally considered 16 images for the present thesis focused on grain matrix-segmentation. However, all the images are likely to be useful for different types of future work in geosciences. We hope that this dataset that we are now going to make public (<https://github.com/Rajdeep002/grain-matrix-dataset>), will serve in a long way to act as an important benchmark in automated image analysis of sedimentary rocks.



## 3.2 Dataset preparation

Since in literature or real life there are no presently existing dataset available, we have prepared a dataset with ground truths for nineteen photomicrographs of clastic sedimentary rocks along with their ground truths. The images of clastic sedimentary rocks (mostly sandstone) observed under a petrographic microscope were used as the source for the image processing work. In order to study the rocks under a petrographic microscope, one needs a very thin slice of the rock sample, typically with a thickness of the order of 0.03 mm, the thickness at which the minerals constituting the rock becomes transparent and all its optical properties are revealed under the microscope. Usually geologists collect these rock samples from the remote field areas, bring them to the laboratory, use rock cutters to cut a slice of a few mm, and mount that on a glass slide. Then this thin glass mounted rock slice is polished on the rotating lap using fine-grained corundum powder, to bring it to the desired thickness of 0.03 mm and covered with a coverslip.

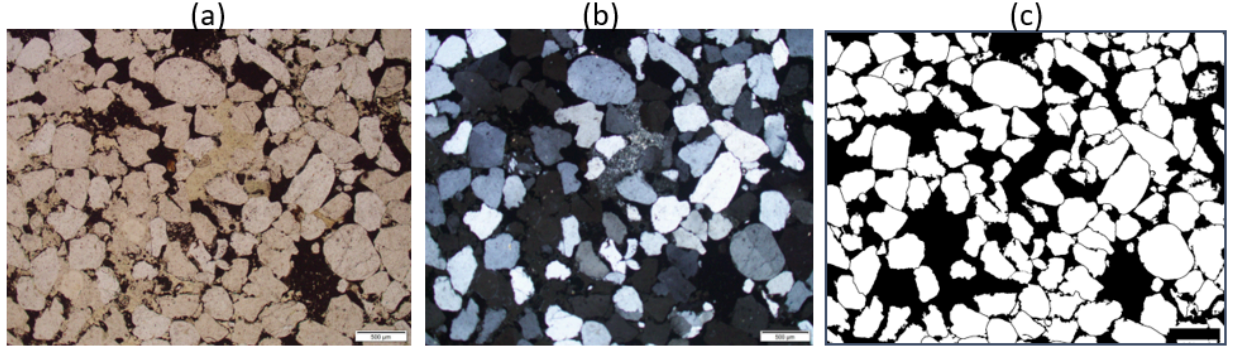
The petrographic microscopes used by the geologists (Fig. 3.1) have two polarizers fitted to it. The lower one is called polarizer whereas the upper one is called analyzer. The translucent thin section is placed over a polarized incident light, and after magnifying it by objective lenses, we can see the image of the rock thin section that comprises fragments of different mineral and rock fragments. This view of the slide observed under a single polarizer (below the stage on which the thin section is placed) is called a plane polarized view or PPL view. So the ordinary light is transformed into polarized light before it enters the thin section. In PPL view, one can see all the detrital grains with their natural body color, as well the matrix. The refractive index of the minerals, their boundaries or their pleochroic color variation is well-displayed.



**Fig. 3.1:** Outline and parts of a petrographic microscope

However, when the analyser (Fig. 3.1) is also inserted we see a set of different colors known as the birefringence colors. In general isotropic materials like glass etc. remain dark irrespective of the microscope. Anisotropic materials on the other hand are of two types: with uniaxial crystal and biaxial crystal. Depending on the relative direction of cutting of the respective crystal with respect to its absolute orientation, numerous kinds of colors are generated at each crystal point on the thin section. As the polarized light passes through the rock thin section, light again splits up in two rays that have different speed and vibration direction. The analyser kept at the top (Fig. 3.1) again allows the light to vibrate only in one direction. Thus the different vibration directions are forced to vibrate in a single plane allowed by the analyser, and due to the interference of two rays the birefringence color is produced. The view of the thin section with both polarizer and analyzer inserted, and showing the birefringence color of the minerals, is called a cross polarized view or XPL view. In a cross polarized view, some of the grains become completely dark (called extinction position) due to destructive interference of the two light rays. In our studies, from each rock thin section, a PPL view and an XPL view is taken. The stage (the position of the thin section) remains constant, so that they can be superposed capturing all the optical features of the grains/clasts. The captured images are quite large in size (e.g. 1920 pixel x 2448 pixel, 1536 pixel x 2048 pixel etc). The dataset consists of a total of 19 images, and as has already been explained previously in Chapter 1, sixteen (16) images have been considered for the research work performed in this thesis. All these 16 images are a set of triplets consisting of the

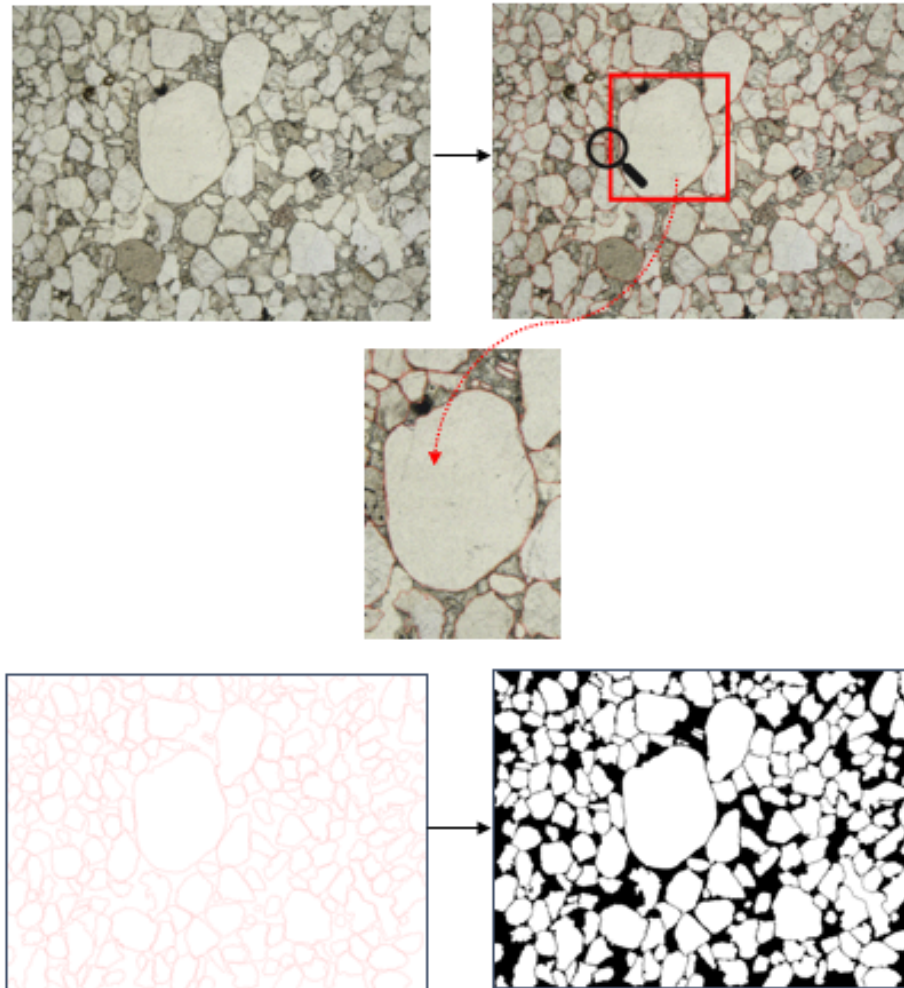
Plane-polar (PPL) view, the Cross-polar (XPL) view and the binary ground truth (GT) image, which we have prepared manually (Fig. 3.2). In the GT image (Fig. 3.2c), the white color indicates grain regions and black color indicates the matrix regions in respective images.



**Fig. 3.2:** (a) PPL image (b) XPL Image (c) GT image.

For reasons explained above, most of the works in sandstone image segmentation rely upon the XPL views only, and our first works (presented in Chapter 3, Chapter 4 and Chapter 5) that either involve some classical unsupervised machine learning methods or try to mimic the Geologists' eye through unsupervised models of visual psychophysics, are done on XPL images. However, in the deep learning based segmentation works (Chapter 6), both the types (i.e. both PPL and XPL), have been used via a combined view image by averaging the XPL with its corresponding PPL. This is a unique and novel methodology that we have applied that has also led to an augmentation of the image dataset which therefore comprises the PPL, the XPL, and additionally, the average (PPL, XPL) images.

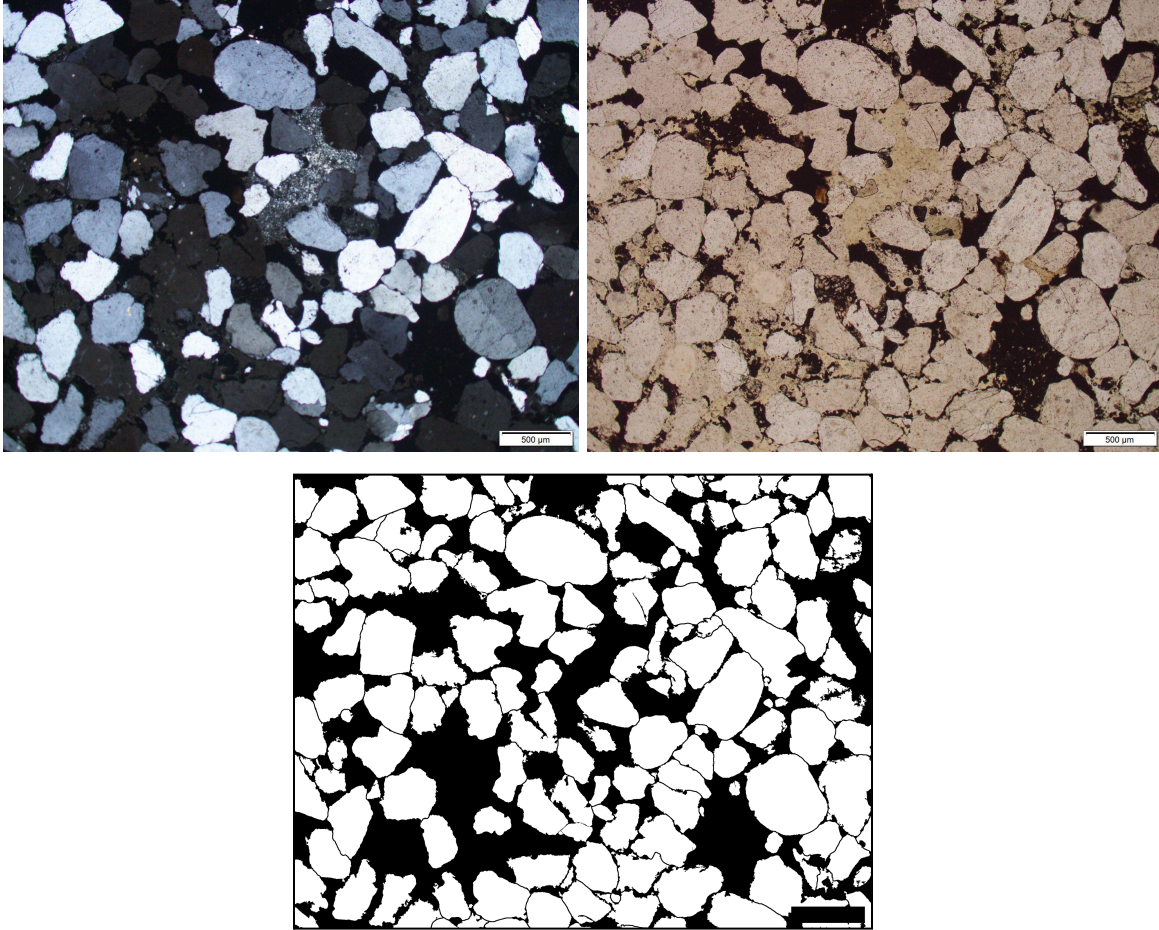
### 3.3 Methodology for preparing ground truth



**Fig. 3.3:** The steps followed in preparing the ground truth image. Here the PPL image has been used for marking the grain boundaries manually.

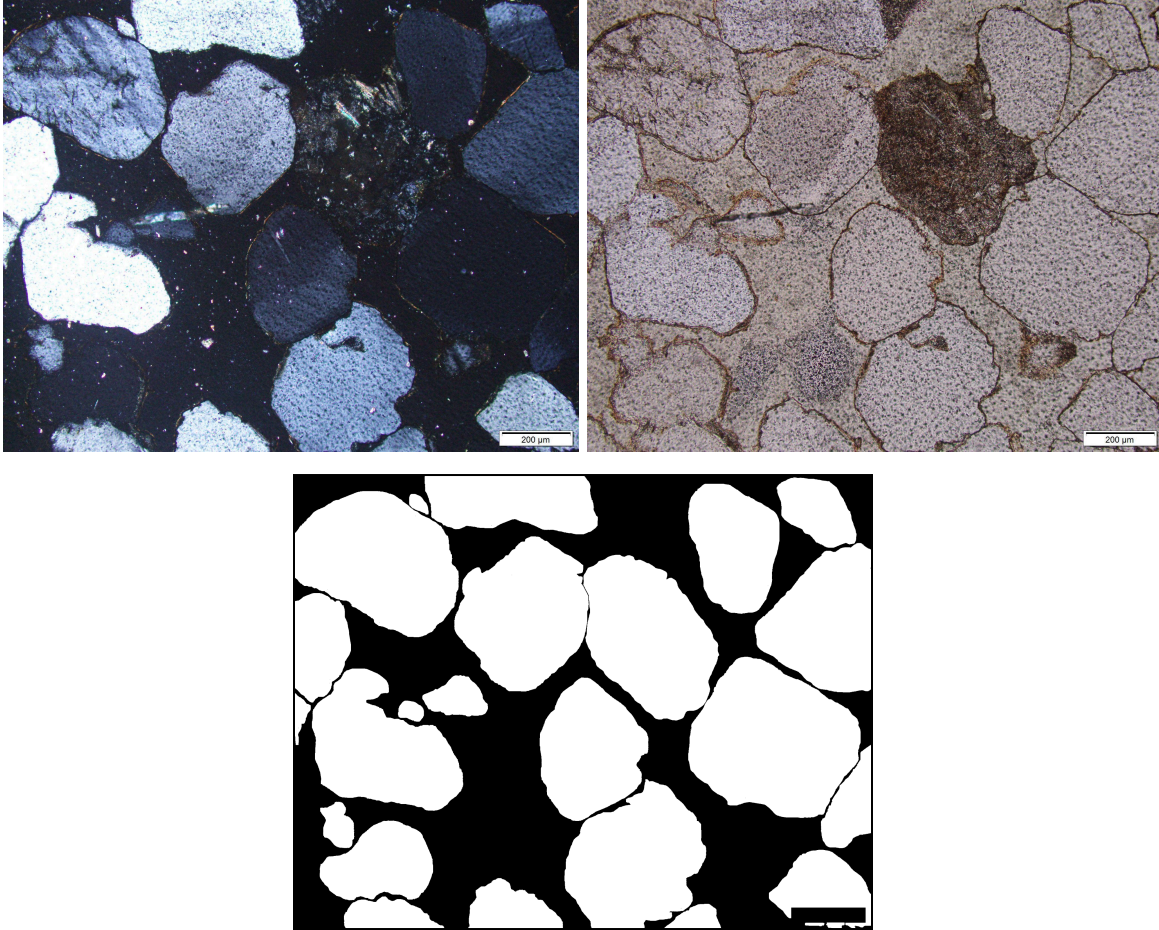
The ground truth preparation is a tedious process that has been outlined in Fig. 3.3. One needs to interact with the geologist with multiple iterations all through this process to correctly fix the grains and the matrix region. Below we demonstrate some of the images (in triplet) from our dataset from Fig. 3.4 to 3.7 that demonstrates the different variations of photomicrographs with respect to size and other characteristics of the grains in our dataset, unlike the images dealt with in many of the publications earlier mentioned. Image 3.8 is a special case outside our dataset that contains some gray areas in its GT image which are difficult to identify as either grain or matrix. In the last (concluding) chapter of this thesis, we have mentioned a possible future work in this

direction using Generative Adversarial Network (GAN) to learn such ambiguous photomicrographs as well. These images we have not used in the present grain-matrix two-class segmentation problem.



**Fig. 3.4:** A sample photomicrograph from our dataset showing the XPL, PPL views and the GT image containing many grains.

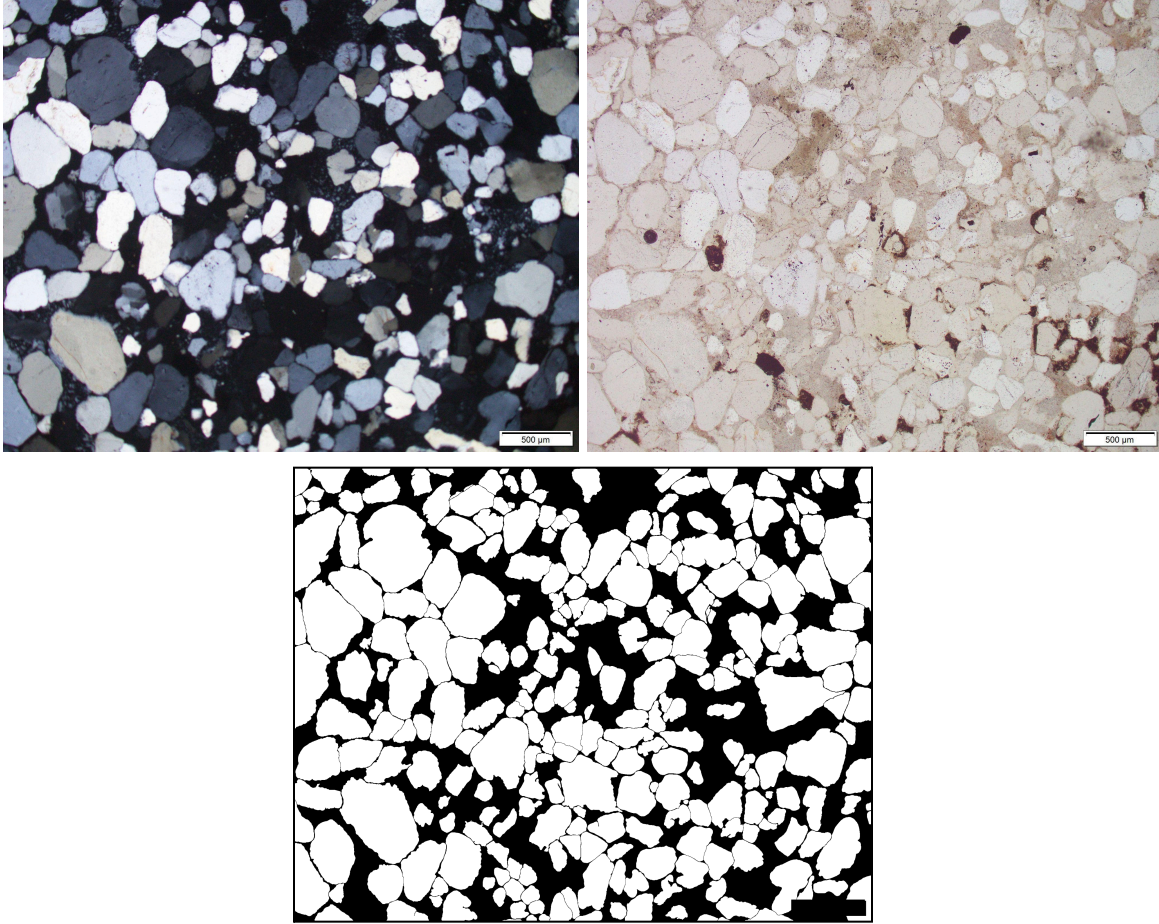




**Fig. 3.5:** A sample photomicrograph from our dataset containing comparatively larger grains, and a considerable amount of matrix area unlike most of the images found in literature.

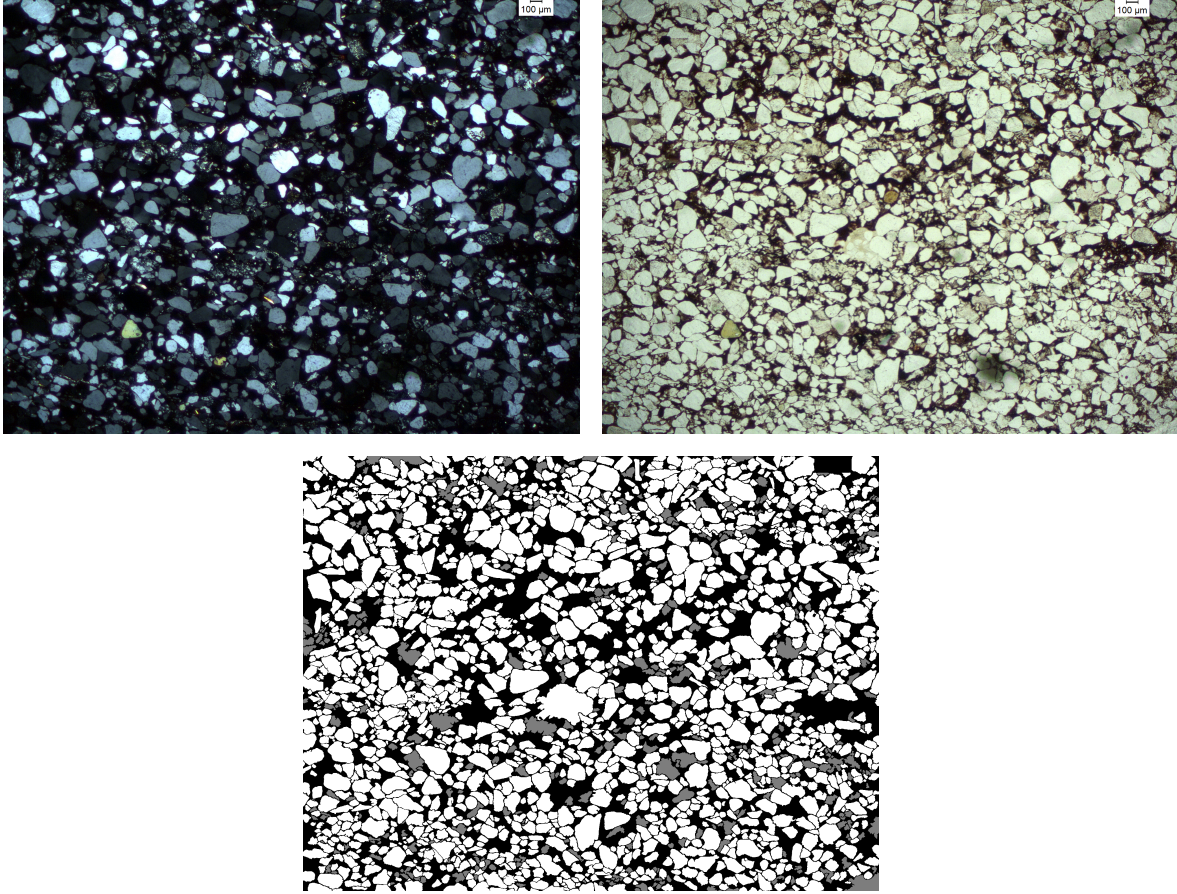


**Fig. 3.6:** Another sample image from our dataset containing larger grains and ambiguous border regions.



**Fig. 3.7:** A sample image from our dataset containing a large number of small sized grains.





**Fig. 3.8:** A sample image from our dataset containing a large number of very small grains. It can be seen that in this image, apart from black and white there are some gray portions in the GT image demonstrating those regions that cannot be marked either as grain or matrix and are hence identified as ambiguous regions in gray color.

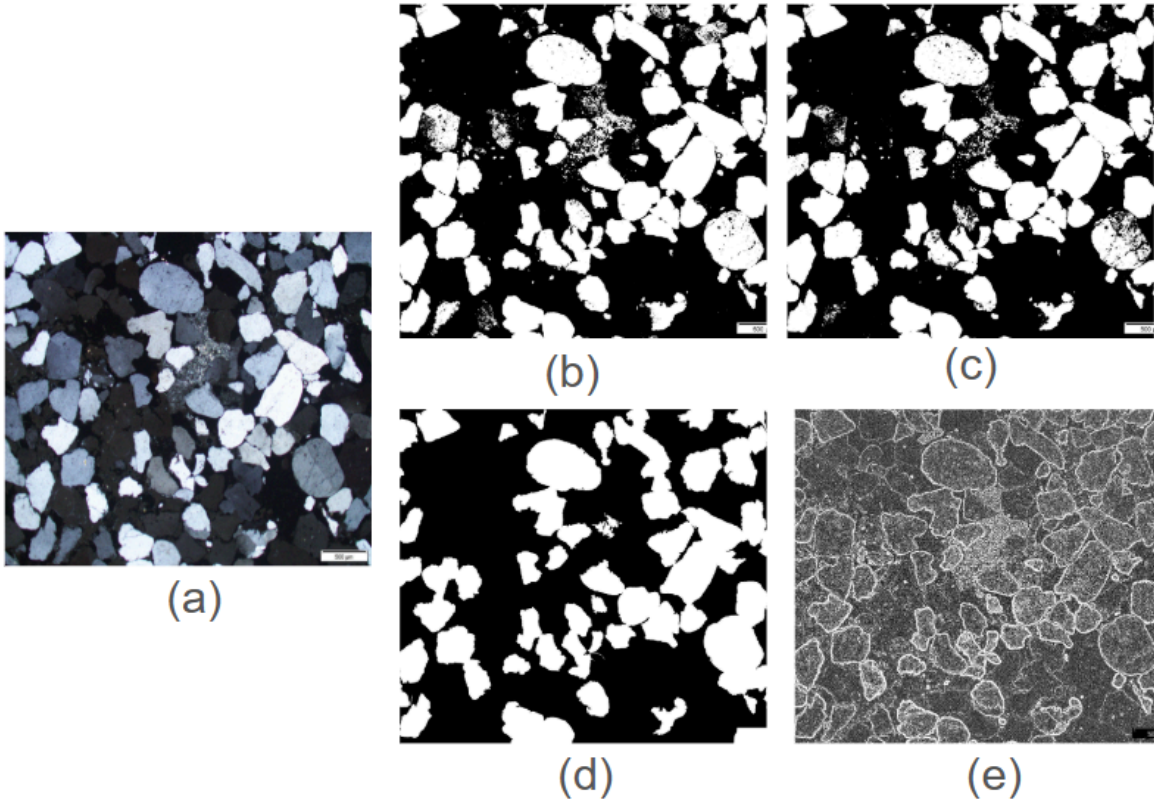
### 3.4 Results based on unsupervised segmentation

We have tested the dataset mentioned above with some well-known machine learning and computer vision based unsupervised techniques. Table 3.2 describes the results under unsupervised segmentation. Among the cross-polar images, Kapur's entropy method demonstrates superior performance in terms of recall, dice score, and Jaccard index, establishing it as the most dependable approach overall. The NCut algorithm demonstrates the highest level of accuracy. Otsu's algorithm achieves the highest precision score. The LBP technique has the

poorest performance across all parameters. Figure 3.9 shows the segmented results of these four unsupervised algorithms.

| Method          | Quantitative Score on Test Images |             |             |             |             |
|-----------------|-----------------------------------|-------------|-------------|-------------|-------------|
|                 | Accuracy↑                         | Recall↑     | Precision↑  | Dice↑       | JI↑         |
| Otsu            | 0.65                              | 0.48        | <b>0.93</b> | 0.63        | 0.46        |
| Kapur's entropy | 0.64                              | <b>0.51</b> | 0.90        | <b>0.65</b> | <b>0.47</b> |
| NCut            | <b>0.66</b>                       | 0.45        | 0.90        | 0.40        | 0.44        |
| LBP             | 0.50                              | 0.48        | 0.61        | 0.53        | 0.37        |

**Table 3.2:** Performance comparison of various unsupervised methods.



**Fig. 3.9:** Output images under unsupervised method. (a) original cross-polar image, (b) Otsu's segmentation, (c) Kapur's entropy segmentation, (d) NCut segmentation, (e) LBP segmentation.

### **3.5 Conclusion**

To the best of our knowledge, there exists no existing dataset for grain-matrix segmentation in clastic sedimentary rocks in spite of the need for automatic segmentation. None of the authors of the published papers in the domain so far have made their dataset public. In fact, in most of the published works, it has been found that the algorithms are tested on only three to four images which too are often found to be quite trivial with respect to the complex tasks in real-life situations. In this regard, the preparation of the dataset of 19 sandstone photomicrographs with wide variations in size and characteristics of grain, and matrix-richness, which will now be made public, is one of the contributions of this thesis that is going to serve future research in the application of computer vision in the analysis of sandstone photomicrographs. We have also attempted several unsupervised machine learning methods to segment these images but the results are not very satisfactory and we need to approach the problem from different angles which will be evident in our next few chapters.

# Chapter 4

## Automatic grain segmentation in cross-polarized photomicrographs of sedimentary rocks using a psychophysics inspired neuro-visual model

### ***Publications:***

1) Das, R., Shankar, B. U., Chakraborty, T., & Ghosh, K. (2021). Automatic grain segmentation in cross-polarized photomicrographs of sedimentary rocks using psychophysics inspired models. *Innovations in Systems and Software Engineering*, 17(2), 167-183.

2) Das, R., Shankar, B.U., Chakraborty, T., Ghosh, K., 2020. Addressing Grain-Matrix Differentiation in Sedimentary Rock Photomicrographs in the Light of Brightness Perception Modeling. In: Mandal J., Mukhopadhyay S. (eds) *Proceedings of the Global AI Congress 2019. Advances in Intelligent Systems and Computing*, vol 1112. Springer, Singapore.

### **4.1 Introduction**

In the first two chapters we discussed in brief a general background of our current problem domain, whereas in the previous (third) chapter, we described the preparation of the dataset in order to design and test models and algorithms to solve the problem in absence of any such. In the current chapter we shall discuss and show that a psychophysics based computational model when applied on the cross-polar images, mimics potentially the visualization mechanism of expert geologists while inspecting and detecting the grains in the background of matrix or cement. At the same time we shall show also how our proposed psychophysics inspired computational mode enhances and pre-processes the XPL images for better capturing of low-level features like edge etc., so that they may be recognized with normal segmentation

methods like clustering based ones, very efficiently. Let us first discuss and be familiarized about the motivation behind our current work which is narrated below.

As we have already discussed in chapter 1, sedimentary rocks that cover almost sixty-five percent of the exposed earth's surface, are deposited layer by layer, forming what is usually referred to as bedding. The detritus formed through weathering of pre-existing rock material are entrained, transported and deposited by glaciers, rivers, wind or ocean waves to form clastic sedimentary rocks. The succession of sedimentary rock strata is critically indicative of the surface processes and climatic conditions prevailing in the past. Evolution of life can also be studied by inspecting sedimentary rocks containing the fossil records from the life forms in historic, and prehistoric times. The segmentation of grains in the cross-polarized photomicrographs of sedimentary rocks is a crucial first step in this study. [Dias et al., 2020] have recently shown that since geological materials are inherently highly variable on a small scale, capturing good-quality images, particularly of the fine details present in the mineral grains that compose the rock, is a major difficulty encountered when a thin section is examined under a petrographic microscope. They combined petrographic concepts and digital image processing methods like embossed surface, negative image, and edge detection techniques to arrive at a practical software based approach to digital image treatment of the photomicrographs. The aim of the present work is also to follow similar trails like that of [Dias et al., 2020] which involves basic calibration of color, brightness, and contrast, but this we attempt to do through the use of a different set of image processing and pattern recognition tools in the spatial domain. The grains are actually the larger rock fragments (Fig. 4.1) to be segregated from the matrix of the rocks which is nothing but the finer-grained sedimentary material, such as clay or silt that surrounds or occurs in the interstices of larger grains. The identification of the grains is usually done by the geologist on microscopic images captured under a polarizing microscope. Any such image may contain hundreds of mineral grains, and manual segmentation is extremely time consuming and tedious, apart from being dependent upon subjective experiences. These problems have recently been discussed in detail by [Mokhles et al., 2019] demonstrating that not only point counting is time consuming and requires extensive efforts from the geologist/petrographer, but its results are also highly subjective and heavily dependent on the petrographer's knowledge and expertise. In this collaborative research work, from the interaction between the expert Geologist and the computer scientist, the requirement was felt of such computer vision algorithm/s that, akin to the

trained geologists' eye, do not get deceived by the wide variation in brightness information [Kingdom, 2011] as elucidated by citing difficult examples in Fig. 4.2. The role of visual psychophysics in brightness modeling and solving the grain-matrix segregation problem in real life situations, was thus felt.

#### *4.1.1 The motivation behind brightness perception modeling for photomicrograph analysis*

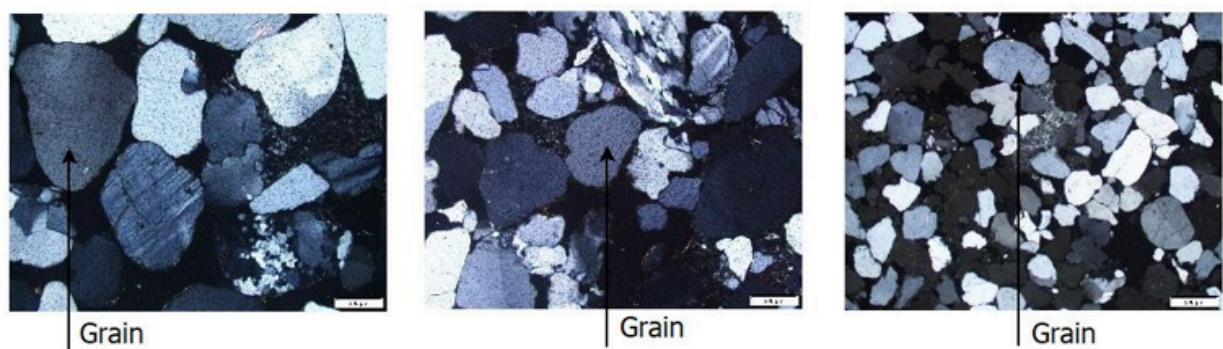
To understand the degree of challenge, with respect to brightness-contrast modeling, it is important to know how color and intensity are generated in the petrographic photomicrographs. The color and intensity of a particular point in a rock thin-section under polarizing microscope depends on the optical property of the crystal at that point, its orientation and optical phenomena viz. refraction, dispersion, absorption, and birefringence, that may occur at that point. The petrographic microscope consists of two Nicol prisms having their polarizing planes oriented perpendicular to one another. When only an isotropic material such as air, water, or glass exists between the polarizers, all the light is blocked. However, most crystalline materials and minerals change the vibration direction of light passing through them, allowing some of the altered light to pass. Using one polarizer makes it possible to observe certain optical properties (e.g. refractive index of the minerals) in plane polarized light; using two polarizers allows observation of certain other properties (e.g. birefringence) and is referred to as the cross-polarized view. A look at a cross-polar image captured through the microscope (Fig. 4.2) and its corresponding ground truth, reveals even to a non-expert the motivation for using computer vision algorithms on one hand, and brightness-contrast psychophysics modeling, on the other. Depending upon the orientation and type of the various crystals present in such slides, which a geologist is interested in extracting, the color and intensity of the different segments of the picture vary widely from very dark to very bright although such apparent variation may not actually be meaningful, and hence may mislead. For instance, this is especially true in the matter of identifying the larger rock fragments called the grains (see Fig. 4.1 where one finds, alongside the grains, the matrix of the sedimentary rocks i.e. the finer-grained sedimentary material, such as clay or silt that surrounds or occurs in the interstices of larger grains). To segment these out, for example to count such grains through an automated algorithm, we need computer vision algorithm/s that, akin to the



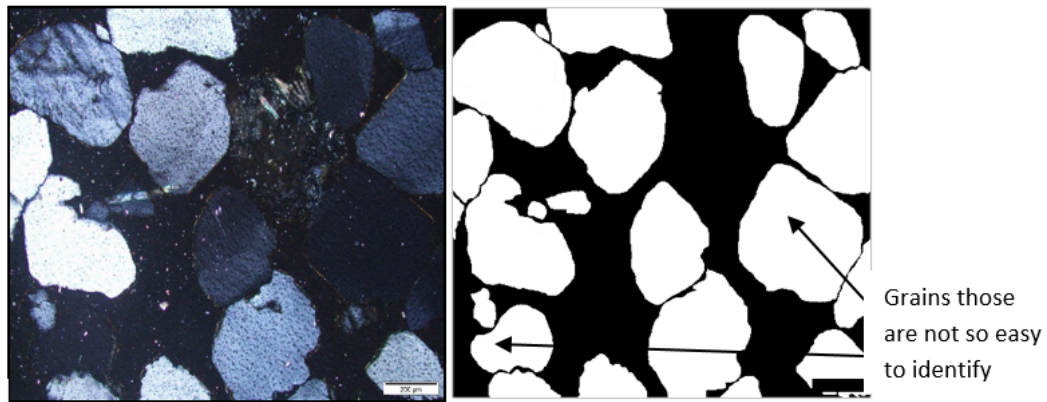
trained geologists' eye, do not get deceived by the wide variation in brightness information [Kingdom, 2011] as elucidated by citing difficult examples in Fig. 4.2. Through appropriate image processing methods, such segmentation can be achieved efficiently and accurately whereas for human beings it can be a time consuming and tedious job [Mokhles et al., 2019]. In other words, the automatic detection and analysis of the photomicrographs can be efficient, fast, and accurate provided the brightness modeling is done competently.

#### *4.1.2 The psychophysics of brightness-contrast and its modeling*

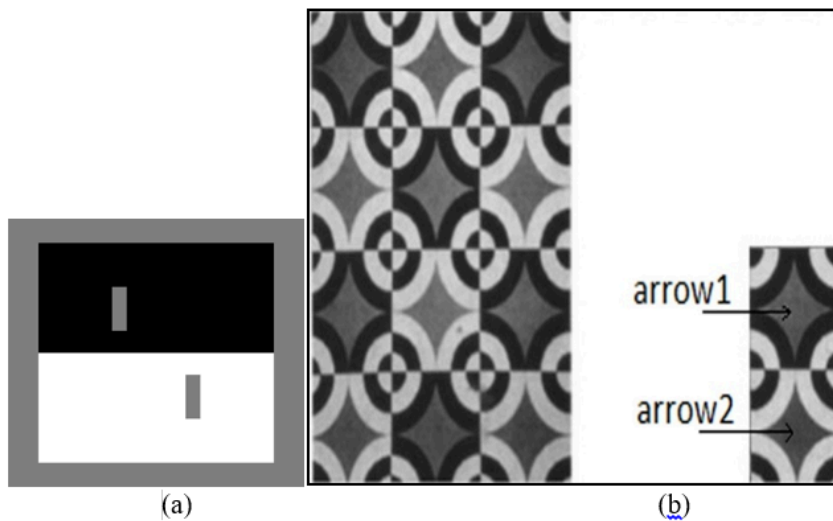
The perceived brightness of any surface depends on the brightness of the surfaces that surround it, a phenomenon termed as brightness induction [Kingdom, 2011]. However, several studies by Helson [Helson, 1963] and also Jameson & Hurvich [Jameson and Hurvich, 1989], reveal that brightness related perception is a complex task in the brain [Kingdom, 2011] as a result of brightness induction occurring from the neighbourhood of any Region of Interest (ROI). The complexity is attributed to the fact that such induction is in no way unidirectional. Rather, brightness contrast and brightness assimilation are two opposite bi-directional aspects of brightness induction occurring in our vision. The former aspect is believed to be exhibited by filtering out the low spatial frequency components of the stimulus and the later aspect by filtering out the high spatial frequency components [Blakeslee et al., 2004].



**Fig. 4.1:** Typical cross-polar photomicrographs from sedimentary rocks displaying the large rock fragments, viz. grains, with the finer-grained sedimentary material, such as clay or silt termed as matrix, occurring in the interstices of the larger grains.



**Fig. 4.2:** A typical sedimentary rock photomicrograph and its ground truth marked by the expert. From a comparison between the two, it is clearly evident that there is a wide variation of brightness of the grains, and a couple of these, marked by arrows, are so dark that they are practically invisible in the cross-polar images.



**Fig. 4.3:** Examples of (a) brightness contrast and (b) brightness assimilation. In (a) the two test patches of equal intensity appear to be of different brightness based on the surroundings in such a way that one on a darker background appears to be of brighter shade than the other one on a lighter background. In (b) the brightness induction is opposite. The surrounding darkness or lightness seems to have been assimilated into the centremost patches (marked by Arrow 1 & Arrow 2) thus representing a phenomenon directionally opposite to contrast.

The two typical examples of brightness contrast and assimilation have been shown in Fig. 4.3 which has earlier been also discussed in Chapter 2 (Fig. 2.9). In Fig. 4.3(a) the two test patches of equal intensity appear to be of different brightness based on the surroundings, such that the



patch on darker background appears brighter than the other one on a lighter background. In 4.3(b) the brightness induction is opposite. For each of the centre most patches (marked by Arrow 1 & Arrow 2), the darkness or lightness of the immediately outer bands has been assimilated into these, thus representing a phenomenon opposite to contrast.

There have been several attempts to computationally model this complex psychophysics of brightness perception [Blakeslee et al., 2004], [Ghosh, 2012] in the recent past, but seldom has that been applied to perform complex brightness perception-based image analysis tasks as in geological photomicrographs or in medical images. The neural networks in the central visual pathway and their models [Ghosh 2016], except for a few general applications related to simple tasks like line detection or more complex tasks like object semantic representation [Wei et al., 2020], [Wei 2016], have rarely been exploited to perform critical classification tasks in image segmentation in spite of the fact that the same is done visually by experts in medical or geological sciences while using varying imaging modalities.

In this chapter we shall try to understand through a neuro-visual modeling approach how the parallel channels, especially the Magnocellular and the Parvocellular channels in the central visual pathway can play a significant role in performing automated image segmentation of the cross-polar photomicrographs, designing three corresponding filters to perform computer vision tasks. These filters mimic three modes in which, we propose, an expert's eye segments the grains under a petrographic microscope viz. vision at a glance (VG), vision in strong contrast mode (VSC), and vision-with-scrutiny (VS), slightly modifying the proposal of Hochstein & Ahissar [Hochstein and Ahissar 2002]. To this end we examine a few cases of optical illusion, related to brightness perception, with our proposed model. We then investigate if this domain knowledge provides clues to a psychophysics inspired automated figure-ground grouping task for the grain and the matrix in cross-polarized photomicrographs. The rest of the work is organized as follows. Section 4.2 is a long Materials and methods section thoroughly explaining and establishing the entire methodology followed. The proposed neural model of brightness perception has first been psychophysically vetted, thus establishing the three filters mimicking the central visual pathway. The original and filtered images are then segmented using an algorithm that combines a modified K-means++ algorithm in CIE Lab color space. This is followed by binary segmentation, optimized through a statistical measure termed  $\beta$  index. Appropriate indices like Dice score,

Jaccard index, and Accuracy are lastly computed for the sake of comparison with respect to the expert vetted ground truth. Section 4.3 provides all the results along with their interpretation. The chapter ends with a Conclusion section indicating the possible future directions of research.

## 4.2 Materials and methods

The rock thin-section images used in the present study have been collected by the geologists of the Indian Statistical Institute. Each of these digital images in our dataset, as also mentioned in Chapter 3, are generated through microscopic study of rock thin sections, which are slices of about 0.03 mm thickness, mounted on glass slides. A chip of rock is, at first, cut out of a sample. One side is then well-polished and carefully glued to a clean glass slide using a transparent adhesive. The rock chip is then finally ground down to the desired final thickness of about 0.03 mm. The dataset of all the digital images have been collected and the corresponding ground truths have been prepared under the guidance of the expert geologist in our group. To mention more specifically, the ground truth image has been built manually under expert supervision by filling the grain portions of a sedimentary rock photomicrograph in white making the rest, i.e. matrix portion black, a process described in detail in Chapter 3.

The proposed methodology of automatic grain segmentation from these photomicrographs is a combination of several component methods that will be explained in separate subsections. A brief outline is as follows:

---

**First**, the cross-polarized images are filtered through the psychophysics inspired filter. This filter for mid-level vision basically proposes that mid-level vision in Human Visual System (HVS) occurs through a two-step process in three possible modes: a) *Vision-at-a-glance* (VG) that uses incoming signals of the Magnocellular (M) pathway, b) *Vision-with-scrutiny* (VS) that uses the previous step to prioritize areas of scrutiny and uses the Parvocellular pathway to scrutinize those areas combining P and M pathway signals (PM), and c) *Vision-in-strong-contrast mode* (VSC) that solely uses the Parvocellular (P) pathway information.

**Second**, the filtered RGB images are converted to the CIE Lab space which unlike RGB and other color spaces is designed to approximate human perception, and the L component of Lab space actually closely matches lightness/brightness perception in HVS. This is a strong motivation for converting the psycho-visually designed filter outputs to this color space.

**Third**, these images are then clustered into 20 clusters using a K-means++ algorithm to closely resemble the original photomicrograph in consultation with the expert.

**Fourth**, a set of sequential operations are performed on the clustered images to arrive only at two final clusters in white and black, representing the grain and the matrix respectively through a process of optimization involving a parameter called *beta index*.

**Fifth and the last**, appropriate indices like Dice score, Jaccard index, Accuracy are computed for this final binary image with respect to its ground truth.

---

These steps are now described in more detail in the following subsections.

#### 4.2.1 The neural model for brightness perception

The proposed neural model has been designed as follows. M & P channels have been modeled using a 3-Gaussian feed-forward neural network model, called the Extended Classical Receptive Field Filter (ECRF) [Ghosh, 2012], [Ghosh, 2016]. In contrast to the 42 Difference of Gaussians in the ODOG filter designed by Blakeslee and McCourt [Blakeslee et al., 2004], The ECRF filter is composed of 3 Gaussian functions of varying widths representing the classical excitatory center, the classical inhibitory surround and the disinhibitory extended surround regions of the visual receptive field. It can be mathematically represented as follows:

$$ECRF(\sigma_1, \sigma_2, \sigma_3, x, y) = A_1 e^{-\frac{r^2}{2\sigma_1^2}} - A_2 e^{-\frac{r^2}{2\sigma_2^2}} + A_3 e^{-\frac{r^2}{2\sigma_3^2}} \quad (1)$$

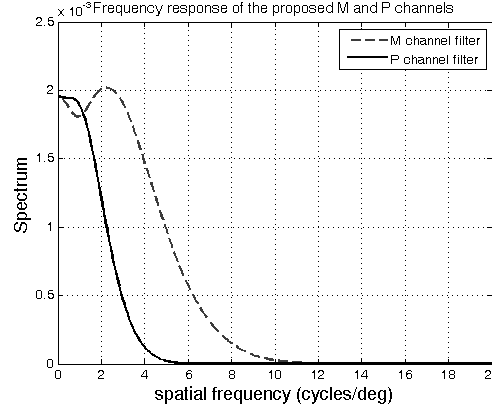
where,  $A_1$ ,  $A_2$ , and  $A_3$  signify the weights to the center, surround and extended surrounds. This neural network model has also recently been considered dynamically by other researchers to make mid-level representations and corresponding applications [Wei et al., 2020], [Wei, 2016]. We further propose that the two spatial filters which model the mid-level neuronal networks as

ECRF described by the above equation, can represent the extended-classical receptive fields of the P and M neurons in mid-level. To substantiate this claim, let us first put forth the corresponding sampling intervals and weights to the center, surround and extended surrounds for these two filters in Table 4.1.

| <i>Filter Parameters</i><br><i>Mid-level Neurons</i> | $A_1$ | $A_2$ | $A_3$ | <i>Sampling interval (s)</i> |
|------------------------------------------------------|-------|-------|-------|------------------------------|
| <b>The Magno neurons (M)</b>                         | 10    | 1     | 0.08  | 0.5                          |
| <b>The Parvo neurons (P)</b>                         | 10    | 0.5   | 0.01  | 0.25                         |

**Table 4.1:** Spatial filters representing the two (M & P) parallel visual channels

Table 4.1 shows that the number of free parameters in the ECRF equation has been reduced from that in Equation (1). The ratio of the three variances of the Gaussians is fixed [Ghosh, 2016]. The variation in sampling density effectively changes the size of the ECRF filter so that at lower sampling density the ECRF of the M filter has lower spatial resolution than the P, as already mentioned for the model based on biological findings. In the value of parameters, it may be noted that inhibitory surround's contribution in the model is double for M as compared to P like the sampling interval also, while the disinhibition is eight-fold more.



**Fig. 4.4:** Frequency response of the proposed P and M filters. The proposed M channel filter (dotted line) exhibits an intermediate kink and a slower fall-off compared to the P (solid line). This M channel behavior is found to bear similarity to the mid-level neuronal cells from lesion studies. The P channel exhibits similar characteristics as shown by mid-level P neurons.

Let us next take a glance at the frequency responses of these two filters when expressed in one-dimension (Fig. 4.4). We find that they bear definite analogy, however crude, to the nature of the spatial frequency dependent contrast sensitivity functions (CSF) in the central visual pathway as obtained from physiological and behavioral experiments. The solid curve in Fig. 4.4, whose corresponding spatial filter represents the lower row in Table 4.1, is analogous to the “Parvo alone” contrast sensitivity function curve experimentally obtained by lesion studies conducted in LGN on macaque monkeys (*Macaca nemestrina*) [Merigan et al., 1993]. In the present work, we suggest that this “Parvo alone” mechanism actually represents “vision-in-strong-contrast” mode (VSC) which is a mechanism often resorted to by an expert performing a very careful inspection and under great attention in medical, or geological (as in present case) image interpretation. Though the analogy is crude, it should still be borne in mind that neurobiological responses are different under suprathreshold conditions [Bowker, 1983]. We are merely trying to compare the nature of the curves. Similarly, the dotted curve in Fig. 4.4, whose corresponding spatial filter represents the upper row in Table 1, has an intermediate kink followed by a similar slower fall-off as in the “Magno alone” CSF curve of Merigan and Maunsell [Merigan et al., 1993], though in the Merigan and Maunsell curve the kink is less prominent. This “Magno alone” filter, it is proposed, and justified below, represents the “vision-at-a-glance” (VG) mode which, though may have some significance in day to day life, is likely to assume less importance when it comes

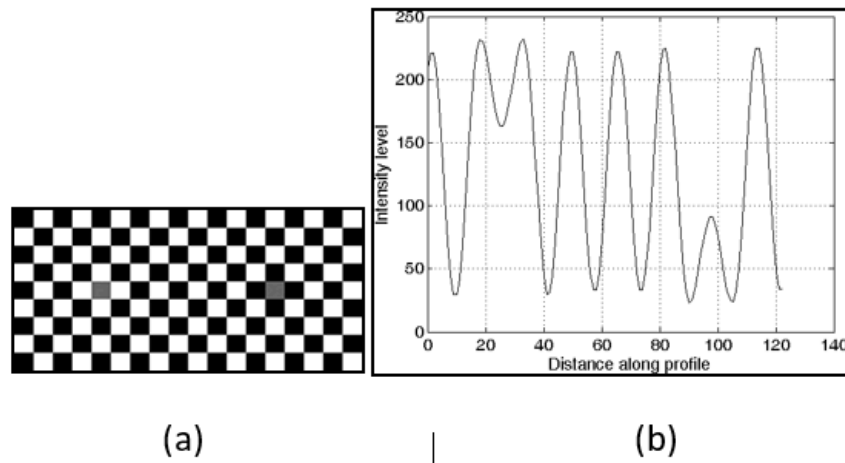
to image analysis type of works performed by the experts. Hence, we also propose a third filter for which the P filter with higher spatial resolution operates on the output of VG, i.e. “M alone” filter to represent a mode of vision (PM combined) that we refer to as “vision-with-scrutiny” (VS), using the concept of Hochstein and Ahissar [Hochstein and Ahissar, 2002]. This mode, we suggest, is also sometimes resorted to by the expert performing inspection, e.g. in our present domain of photomicrograph analysis.

Based on the latencies of visual responses of neurons in different cortical areas, Maunsell et al. [Maunsell et al., 1990] and Bullier [Bullier, 2001] argue that such characteristics of the M channel like faster conduction, high contrast sensitivity, poor chromatic selectivity, larger receptive fields and lower sampling rates makes it a probable candidate for a first-pass “vision-at-a-glance”. The Reverse Hierarchy Theory (RHT) of Hochstein & Ahissar [Hochstein and Ahissar 2002] also envisages this initial “vision-at-a-glance” to include results of automatic and implicit bottom-up processing, which makes the initial explicit perception introspectively direct without conscious antecedents, to be followed up later by more attentive modes of vision like “vision-with-scrutiny” (VS) and “vision-in-strong-contrast” mode (VSC) as proposed here. These modes are to include more indirect conscious perceptual constructs. Before we study the applicability of the suggested VS and the VSC modes of vision of the expert in the image analysis task of photomicrographs, we first try to validate the neural model of the two basic filters viz. “P alone”, and “M alone” representing the two most important parallel channels in the central visual pathway [Merigan et al., 1993], using the context-dependent opposing nature of brightness induction as evident in some related optical illusions in literature. Once we are successful with this psychophysical validation, we shall, as proposed, apply the VS and VSC modes in vision in dealing with the difficult problems in the cross-polar photomicrographs.

#### *4.2.1.1 Psychophysical validation of the neural model using brightness induction illusions*

##### **The checkerboard illusion**

The checkerboard illusion (Fig. 4.5a) of De Valois & De Valois [Valois et al., 1988], is one of the difficult problems in brightness perception [Kingdom 2011] since it is known to produce two opposite brightness effects (viz. assimilation and contrast) in high and low spatial frequency of the inducing gratings respectively [Blakeslee et al., 2004] (Fig. 4.5a and 4.6a). We shall now try to validate the neural model presented above in the light of such intricate as well as “simpler” psychophysical stimuli.



**Fig. 4.5:** (a) The Checkerboard stimulus at 8 pixels width per column (b) Horizontal intensity profile through the two test patches of the output image filtered by P channel, demonstrating brightness assimilation from the surrounding, with the left patch appearing brighter, as also observed in psychophysical experiments.

To this end, the neural model proposed above is now further systematized in the light of psychophysics. We propose that brightness perception being an integral part of the mechanism of attentive vision in the visual pathway, should involve a final choice of channel or combination thereof, in the mid-level vision, possibly through the cortico-geniculate pathway. Such choice is likely to be based on the initial information extracted in the visual cortex through the VG or “M alone” mode, and subsequent feedback to the mid-level after computing certain features from this preliminary output. To this effect, we propose an algorithm for attentive/scrutinized vision:

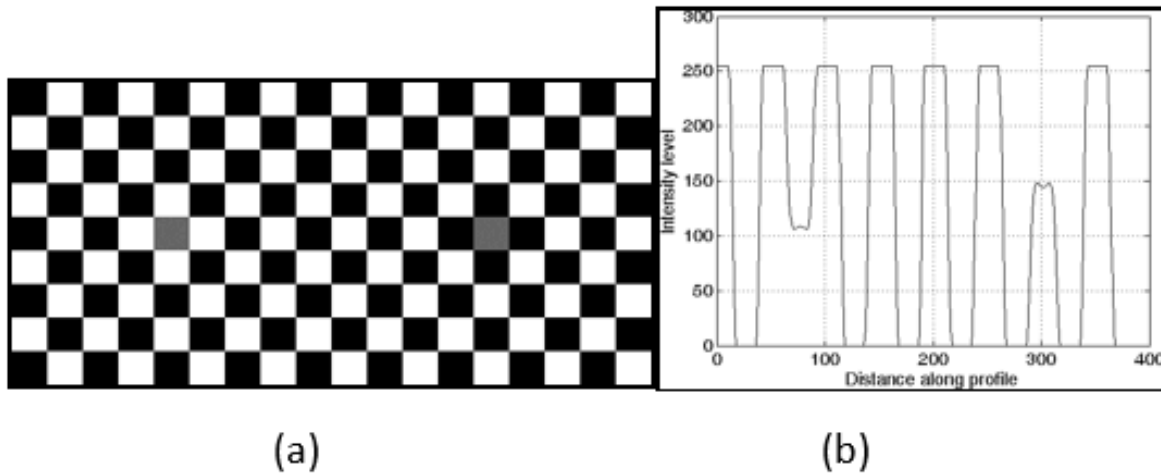
#### Algorithm 4.1: Visual filter selection

|                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <p><b>If:</b><br/>the initial M channel output identifies that the background around the test patches is <i>uniform</i> by virtue of satisfying <b>condition 1</b><br/>OR<br/><b>condition 2</b></p> <p><b>then:</b><br/><i>Either P channel is invoked, and the brightness percept is formed by P alone</i><br/><b>(Here, the VSC mode often adopted by experts in difficult image analysis tasks).</b><br/>OR<br/><i>P channel is invoked and operated on the initial M output to produce the final percept PM</i><br/><b>(Here, the VS mode that may also sometimes be adopted by domain-specific experts)</b></p> <p><b>Else:</b><br/><i>M channel alone produces the brightness percept</i><br/><b>(Here, the VG mode that is generally not applicable for expertise-based image analysis, but may have some implication in everyday life)</b></p> | <p><b>Remarks:</b> Uniform background means M channel identifies</p> <p><b>Condition 1:</b> <i>uniformity with respect to absence of sharp intensity changes in the background outside the test patch, i.e. if the vertical and horizontal line profiles of the background above, below and on either side of the test patch do not contain contrast edges, it will be accepted as uniform background.</i></p> <p><b>Condition 2:</b> <i>uniformity with respect to the direction of Weber contrast* across all the edges of the test patch, i.e. if such contrast of the test patch with its background is unidirectional (has the same sign) both across its horizontal and vertical edges, we call it high contrast surrounding or uniform background in both directions.</i></p> <p>*Weber contrast = <math>(I - I_b) / I_b</math>, with <math>I</math> and <math>I_b</math>, representing the luminance of the region of attention (test patch) and the background luminance, respectively.</p> |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|

This algorithm implies that when any one of *condition 1* or *condition 2* is obeyed by the initial M output, then P channel is invoked and when either both the conditions are violated or at least one is violated and the remaining is inconclusive, then the visual system may continue mostly with the M channel output. For instance, it is easy to see from Fig. 4.7b that the M output for any checkerboard stimulus (Fig. 4.7a) produces the sharp adjacent undershoots and overshoots and



hence *Condition 1* is clearly violated. Alongside, it is also easy to see that the Weber contrasts calculated across both horizontal and vertical edges of each test patch in the M filtered output are same, i.e. *Condition 2* is obeyed and P channel is thus preferred. The experimental findings [Blakeslee et al., 2004] are vindicated by the P outputs in Fig. 4.5b and Fig. 4.6b respectively. The proposed algorithm is further justified in the light of the fact that the same channel produces opposite perception for such a difference in spatial frequency as observed in experiments.

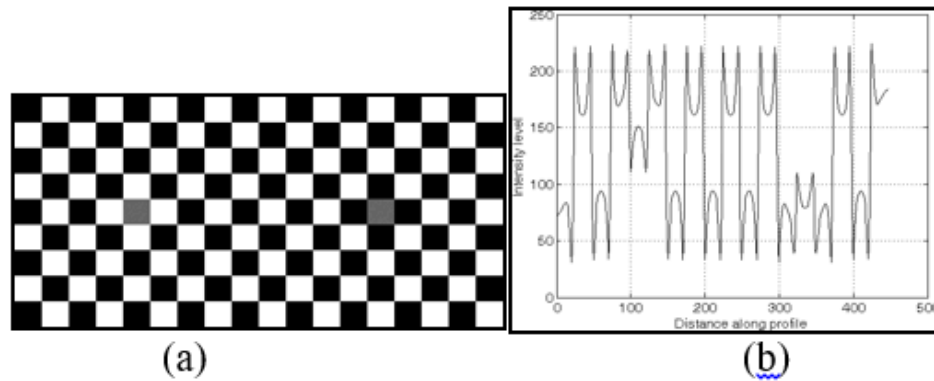


**Fig. 4.6:** (a) The Checkerboard stimulus at 25 pixels checker dimension (b) Horizontal intensity profile through the two test patches of the output image spatially filtered by P channel demonstrating brightness contrast with respect to surround, with the left patch appearing darker, as also observed in psychophysical experiments.

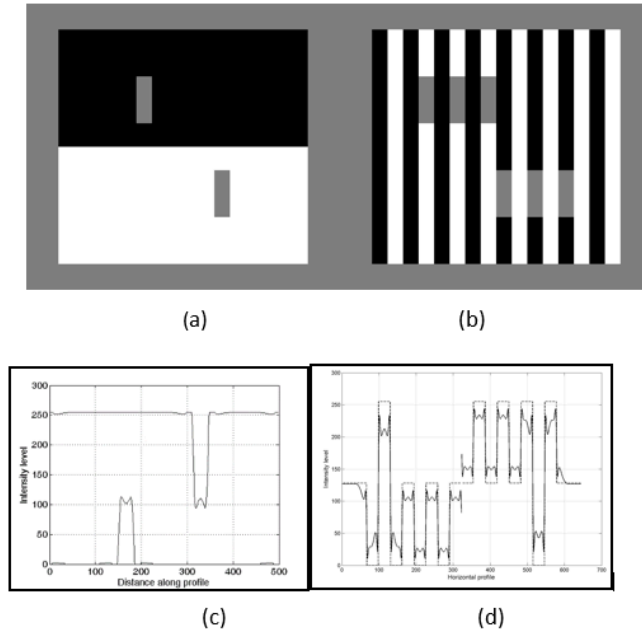
### The SBC and White's illusion

In Fig. 4.8, we present the most common and well-studied brightness perception illusions, providing unified explanations of which have often been considered as benchmark in psychophysics modeling [Blakeslee et al., 2004], [Ghosh 2012]. Fig. 4.8a, represents the classical Simultaneous brightness-contrast effect (SBC) where the same two patches appear to be of different and opposite brightness/lightness as compared to their background and already described earlier in the context of Fig. 4.3a. Here, by virtue of both *Condition 1* (neighborhood is

uniform) and *Condition 2* (Weber contrast is uniform), the P channel is invoked, and the P channel (in VSC mode) correctly predicts the brightness direction as is evident from Fig. 4.8c. The White's illusion in Fig. 8b, presents an exactly opposite percept as compared to the SBC, and the effect of assimilation from surround dominates over contrast, i.e. induction of brightness/lightness from surround takes place in the same and not opposite direction. For the classical White effect, the output(Fig. 4.8d), from M channel filtering displays that both *Condition 1* and *Condition 2* for uniformity test are violated. The violation of *Condition 1* is evident from the adjacent sharp undershoots and overshoots around the test patches (Fig. 4.6d) indicating the presence of contrast edges, while that of *Condition 2* is easy to see since the Weber contrasts across vertical and horizontal edges are of opposite sign. Hence M remains the preferred channel as per the algorithm in section 4.2.1.1.



**Fig. 4.7:** (a) The checkerboard stimulus (b) The corresponding intensity profile of the output image spatially filtered through the proposed M channel displays the sharp undershoots and overshoots in the vicinity of the test patches.



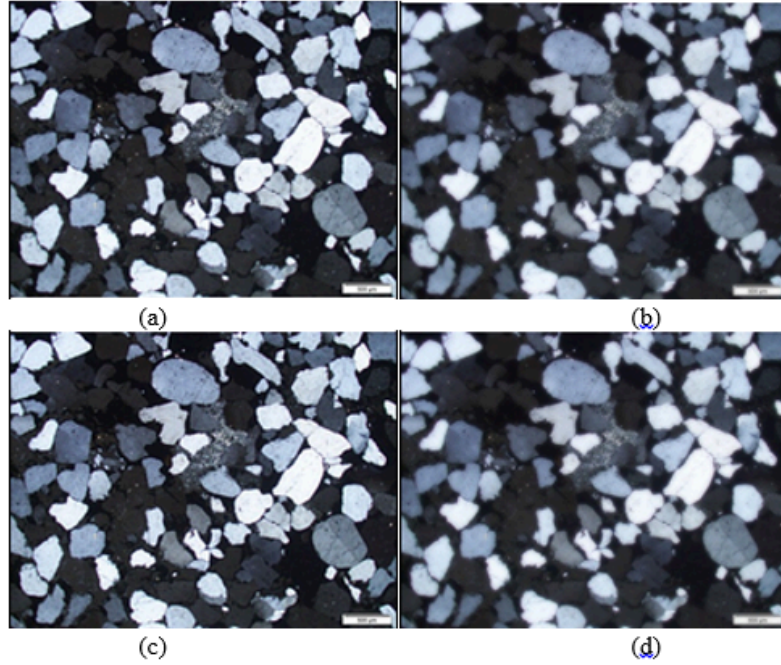
**Fig. 4.8:** (a) The classical example of brightness-contrast. (b) The White effect illusion. (c) Output of P channel over (a). (d) Output of M channel over (b)

Having thus validated the proposed neural model on psychophysics stimuli, we now return to the original typical problem of processing the rock thin-section images in cross-polar mode.

#### *4.2.2 Filtering the cross-polar photomicrographs in the three proposed vision modes*

The digital images are filtered through each of the three filters envisaged to model the three visual modes, viz. VG, VSC, and VS. A sample output is shown in Fig. 4.9, where Fig. 4.9a represents the original image, Fig. 4.9b the output of vision-at-a-glance, Fig. 4.9c the output of vision-in-strong-contrast mode, and Fig. 4.9d the vision-with-scrutiny output. It may be seen visually itself, that vision-in-strong-contrast mode (Fig. 4.9c) extracts some of the grains almost hidden in the matrix in the original image (Fig. 4.9a). The disadvantage of this VSC mode is that unwanted noise, or even tiny silt grains within the matrix may get enhanced and later add to misclassification as framework grain or part of framework sand grain. This is why smoothing/blurring operation is performed in digital image processing. However, a classical blurring operator (e.g. the two-dimensional Gaussian function) will also destroy the grain

outlines. In that respect, it may be seen from Fig. 4.9b that, “vision-at-a-glance” does perform blurring but more or less retains the grain outlines. This is the reason why application of strong-contrast (P filter) on this M output causes image enhancement including edges (Fig. 4.9d) in the “vision-with-scrutiny” mode, and we get a contrast smoothing-enhancement combined output.



**Fig. 4.9:** A sample cross-polar photomicrograph filtered in the three proposed vision modes. (a) The original image. (b) The “vision-at-a-glance” output (c) The “vision-in-strong-contrast” mode (d) The “vision-with-scrutiny”.

#### 4.2.3 Clustering the filtered images in CIE Lab space

The CIE Lab color space is designed to be perceptually uniform with respect to human color vision. It is well-known that this color space is exactly linearly proportional with its three color contents. Especially since the L component of the  $L^*a^*b$  space closely matches lightness/brightness perception in HVS, we have chosen to convert the RGB image to this space in order to carry out the subsequent steps in grain-matrix segmentation. The Lab image is next clustered into 20 clusters by slightly modifying the K-means++ algorithm [Arthur and Vassilvitskii 2006] which is considered as more robust than the classical K-means algorithm

[MacQueen, 1967]. Fränti & Sieranoja [Franti and Sieranoja, 2019] have recently demonstrated that as compared to K-means, the K-means++ leads to more compact as well as better separated clusters. Although the number 20 is a heuristic, it has been verified through various trials, and by taking feedback from the expert that the image with these number of clusters highly resembles the original cross-polar image which the geologist uses for manual grain counting under the petrographic microscope, in terms of capturing all its intensity characteristics. Also, the K-means++ algorithm [Arthur and Vassilvitskii, 2006] has been modified by us with another heuristic toward performing the clustering operation as explained below.

### **The modified K-means++ algorithm**

Before explaining the modification we have made in the K-means++ algorithm, let us first expound the idea behind the K-means++ algorithm [Arthur and Vassilvitskii 2006] itself. The K-means++ algorithm is a randomized variant of the furthest point heuristic. It chooses the first cluster seed as an arbitrarily chosen point in the feature space. Then it chooses a feature point as the next cluster seed with probability proportional to distance between the first seed and itself. In other words, a feature point at furthest distance from the first seed will have maximum probability of being chosen as the next cluster. The third cluster seed is chosen from among the rest of the points with a probability proportional to the greater among the two distances of that point with the two seed points, and so on. This algorithm is an  $O(\log K)$ -approximation to the problem. In our modification, the point at maximum distance from the first seed is selected as the second seed by making the probability of the furthest point being chosen as the second seed equal to unity, thus making sure that no finite probability exists of choosing a point closer to the first seed as the next seed. In case a tie occurs, it is broken arbitrarily. All other steps beyond seed initialization are exactly the same as in the well-known K-means [MacQueen, 1967] or the K-means++ [Arthur and Vassilvitskii 2006] algorithm. The advantages of the K-means++ have been discussed in detail by Fränti&Sieranoja [Franti and Sieranoja 2019] that includes its better performance and higher success rate as compared to most other modifications.

In spite of applying the modified K-means++ algorithm explained above, there still remains the inherent randomness of the K-means algorithm [MacQueen, 1967] even in this clustering

modality, since the first cluster seed is ultimately chosen arbitrarily. Hence, for each image we have run the clustering experiment 20 times for computing the average value of the metrics/indices used for comparison in the final step on the binary segmentation results. However, prior to that there are several other steps that are explained in the subsequent subsection.

#### *4.2.4 Reducing the number of clusters from twenty to two: choice of threshold using $\beta$ -index*

In this section we present all the steps that are executed to arrive at the desired binary segmentation for the sake of comparison with the ground truth where the white pixels represent the framework sand grains and the black ones the matrix that may include many silt grains. The process uses a concept called  $\beta$ -index [Pal et al., 2000] which we first explain below.

##### **Concept of beta index**

Let us assume that a segmentation algorithm results in  $k$  different segments. Let the  $i$ -th ( $i=1,2,3,\dots,k-1,k$ ) region consist of  $n_i$  number of pixels. Let the gray value of  $j$ -th pixel ( $j=1,2,\dots,n_{i-1},n_i$ ) in  $i$ -th region be denoted by  $X_{ij}$  and the mean of  $n_i$  gray values of  $i$ -th region be denoted by  $X_i'$ . Also, if  $n$  be the image size and  $X'$  be the mean gray value of the image, then the  $\beta$ -index is defined as [Pal et al., 2000]:

$$\beta = \frac{\sum_{i=1}^k \sum_{j=1}^{n_i} (X_{ij} - X')^2}{\sum_{i=1}^k \sum_{j=1}^{n_i} (X_{ij} - X_i')^2} \quad (2)$$

It can easily be seen from equation (2) that the numerator of the expression on the right is total variation and the denominator is the within-class variation in the image. So, the numerator is constant but the denominator is variable which is dependent on class homogeneity and number of classes  $k$ . The more the class homogeneity increases, the more will be the  $\beta$ -index. Again, if  $k$  is

1, i.e. the whole segmented region consists of one region only, then the  $\beta$ -index will be equal to 1, and if  $k=L$  (number of gray values in original image), the  $\beta$ -index will assume the value infinity.

### *Steps to binary segmentation of the rock photomicrographs through thresholding*

#### **Algorithm 4.2: Thresholding**

---

- Step1: Let  $C$  be the set of cluster centroids after obtaining the clustered image  $K$ , say, as explained in the previous subsection. We calculate the norm value for each of the 20 clusters' centroids. Then, we sort in ascending order the cluster centroids in  $C$  according to these calculated norm values. Let the sorted sequence be  $C'$ .
- Step2: We compute the  $\beta$ -index values [Pal et al., 2000] for all cluster centroids in  $C'$ , considered respectively as thresholds to segment the image  $K$  into foreground (white) and background (black) regions. Let the sequence of  $\beta$ -index values in same order as in the cluster centroids in  $C'$ , be  $B$ .
- Step3: We find the global maximum  $b$  in  $B$ .
- Step4: Let  $c'$  be the cluster centroid in  $C'$  corresponding to  $b$  in  $B$ .
- Step5: Segment  $K$  into binary image  $S$  by thresholding that image into foreground and background regions using  $c'$  as threshold.
-

#### *4.2.5 Performance measure of final segmentation using Dice score, Accuracy and Jaccard index*

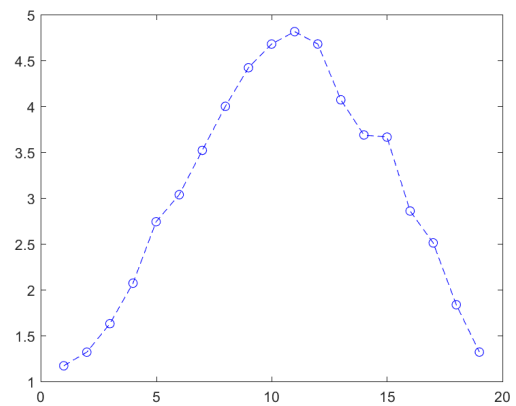
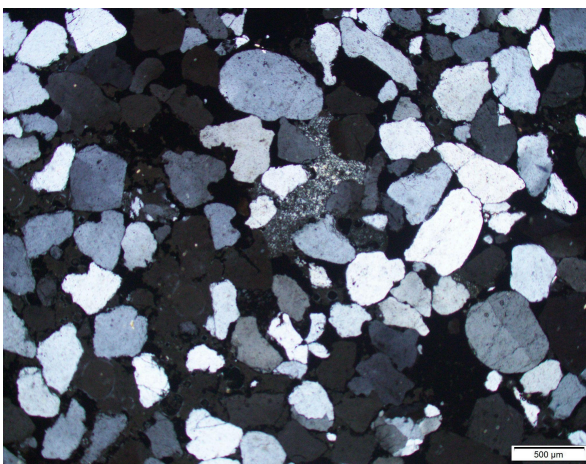
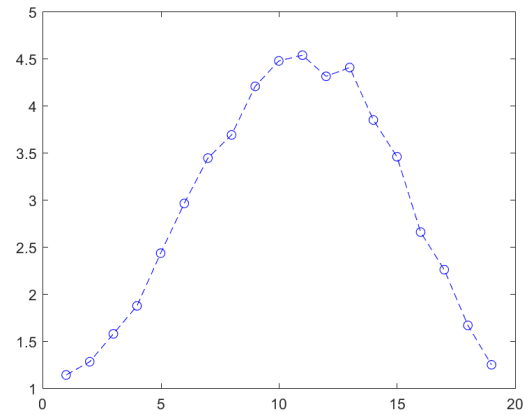
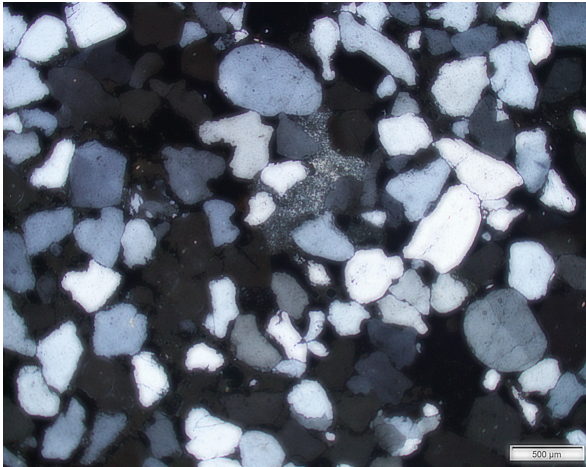
We use the three filtered images obtained through the three vision mode filters, viz. VG, VSC, and VS along with the original image as inputs to the modified K-means++ clustering algorithm and compare the grain-background similarities with ground truth as provided by the expert. We consider that  $X$  is the set of all ground truth grain pixels (in white),  $Y$  is the set of all grain pixels found by digitally segmenting the image (also in white),  $Z$  is the set of all ground truth matrix pixels (in black), and  $W$  is the set of all matrix pixels found by digital segmentation (in black). The performance of each of the filters (viz. P for VSC, M for VG, and PM for VS mode) is compared with the unfiltered original in terms of three indices viz. Dice score, Accuracy and Jaccard index.

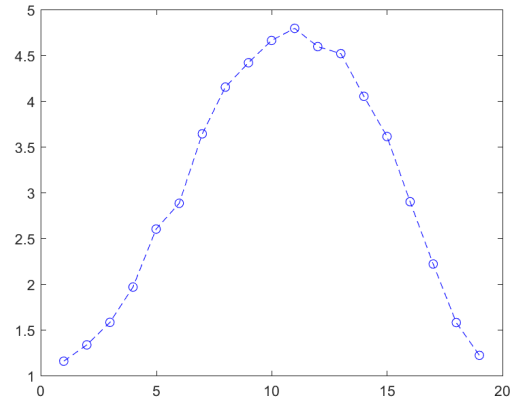
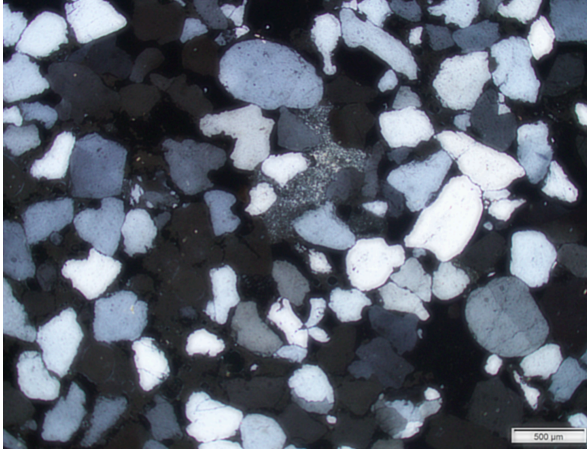
### **4.3 Results and discussion**

Extensive experiments have been carried out on the cross-polar photomicrographs of the sedimentary rock samples collected by the geologists of the Indian Statistical Institute. It was first found [Das et al., 2020] that filters representing all the three modes of vision, viz. VG, VS, and VSC improve upon the performance measures like Dice score, Accuracy and Jaccard index when segmented into black (matrix) and white (grain) as compared to the cases where the original unfiltered image is used. By further modifying the methodology of segmentation, using modified K-means++ algorithm and optimization parameters like the  $\beta$ -index, it has then been found [Das et al., 2021] that “vision-at-a-glance” (M filter alone) is seldom the preferred mode for grain detection through segmentation even for larger grains or when the images have been captured at higher resolution. However, “vision-with-scrutiny” that combines the VG with VSC mode, has often been found to be preferred over the others when dealing with such larger grained or higher resolution photomicrographs giving the highest index measures for these images. On the other hand, for images with many small grains or when captured at lower resolutions, the “vision-in-strong-contrast mode” predominantly provides the highest index values. We present below the results in Fig. 4.10-4.13 of two such representative cases in the three vision modes with  $\beta$ -index plots and also along with corresponding visual comparison with their ground truths

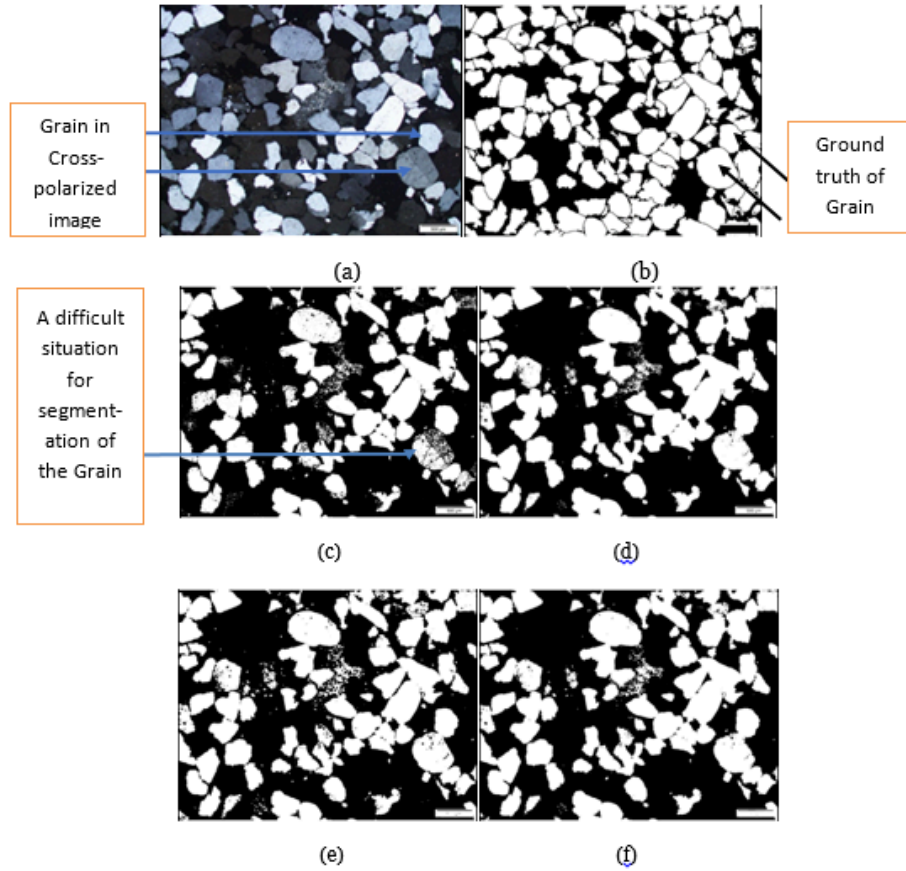


apart from quantitative evaluation with the help of the performance measures in Tables 4.2 and 4.3.





**Fig 4.10:** Results on a typical photomicrograph (Fig 4.9a) displaying its VG, VSC, and VS modes (left column) and their corresponding plots of cluster number (after sorting the clusters in ascending order of their centroid norms) vs.  $\beta$  index plot (right column), for binary segmentation of the image after clustering into 20 clusters and optimizing for 20 runs. 1st Row: left column represents the VG image after convolution by M filter; right column represents the corresponding  $\beta$  index plot. 2nd Row: left column represents the VSC image after convolution by P filter; right column being the corresponding  $\beta$  index plot. 3rd Row: left column represents the VS image after convolution by PM filter and right column, the corresponding  $\beta$  index plot.



**Fig 4.11:** Output images for cross polar image in Fig. 4.9 algorithm 4.2.(a) The same cross-polar photomicrograph (Fig 4.9a Output images for applying on a lower-resolution image show) captured at comparatively lower resolution displaying smaller grains and binary segmentation results of itself and its ground truth marked by expert in (b). The two blue arrows in and the two black arrows in (b) represent a difficult situation for automatic grain segmentation into 2 grains only (i.e. neither 1, because of the very close proximity of the two grains marked in (b), i.e. under-segmentation, nor 3, because of the consideration of the edge existing between two distinct gray levels within the larger grain itself, i.e. over-segmentation. (c) The binary segmented version (optimum in 20 runs) by the mentioned segmentation algorithm on the original image in absence of any of the filters. The marker arrow demonstrates that the edge within the larger grain has been clearly and mistakenly identified. (d) The binary segmented version (optimum in 20 runs) by the mentioned segmentation algorithm on the M filtered image. (e) The segmented version (optimum in 20 runs) by the mentioned segmentation algorithm of the P filtered image (f) Binary segmented version (optimum in 20 runs) by proposed algorithm on the PM filtered image. It can be clearly seen for each of the filtered images in (d), (e), and (f) that the false edge detection with respect to segmentation has been avoided. However, regarding the segregating the two actual grains, the performance of P filter in (e) is better visually. Theoretically also P is the best by achieving the highest values corresponding to the performance measures.

Fig. 4.10 demonstrates the optimum (maximum)  $\beta$  index values where to fix the threshold of the binary segmentation as already explained. Fig. 11 demonstrates visually that for the lower

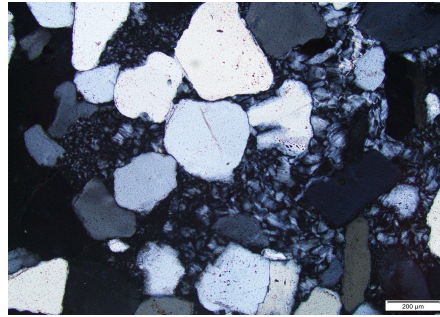
resolution or smaller grain sized images the vision-in-strong-contrast mode (P) filter is the appropriate filter before doing clustering and binary segmentation. The performance measures provided in Table 4.2 also establish the same fact. The psychophysics inspired filters using both K-means++ and the classical K-means clustering may be compared with the classical computer vision segmentation methods discussed in chapter 3. It has been found that the VSC filter with K-means++ clustering consistently outperforms the rest, and the VS filter with K-means++ is consistently the second best in performance and the results are much better as compared to those in Table 3.2 in chapter 3. In other words, the role or inclusion of the P channel filtering is very important for such difficult segmentation tasks where smaller grains in larger numbers are to be extracted from the photomicrographs.

| <b>Vision mode/ Other methods</b>                  | <b>Dice score</b> | <b>Accuracy</b> | <b>Jaccard index</b> |
|----------------------------------------------------|-------------------|-----------------|----------------------|
| Without filter with modified K-means++             | 0.69665           | 0.69840         | 0.53458              |
| At a glance (M filter) with modified K-means++     | 0.74436           | 0.73513         | 0.59301              |
| Strong contrast (P filter) with modified K-means++ | <b>0.75372</b>    | <b>0.74287</b>  | <b>0.60482</b>       |
| With scrutiny (PM filter) with modified K-means++  | <b>0.74623</b>    | <b>0.73613</b>  | <b>0.59530</b>       |
| Without filter with K-means                        | 0.69677           | 0.69809         | 0.53479              |
| At a glance (M filter) with K-means                | 0.73642           | 0.72886         | 0.58349              |
| Strong contrast (P filter) with K-means            | 0.74376           | 0.73442         | 0.59210              |
| With scrutiny (PM filter) with K-means             | 0.74270           | 0.73352         | 0.59073              |

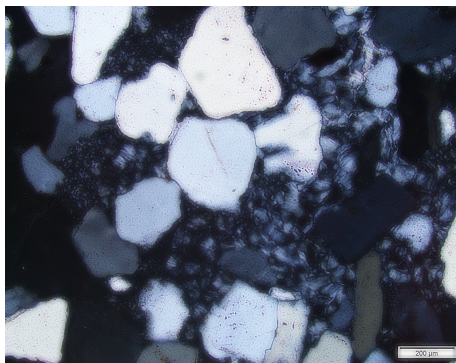
**Table 4.2:** Segmentation indices corresponding to Fig 4.11a for the M, P, and PM filters and comparison with other methods.

We now demonstrate similar results in Fig. 4.12 and Fig. 4.13 for another cross-polar photomicrograph for which the image has been captured at a higher resolution and larger grains

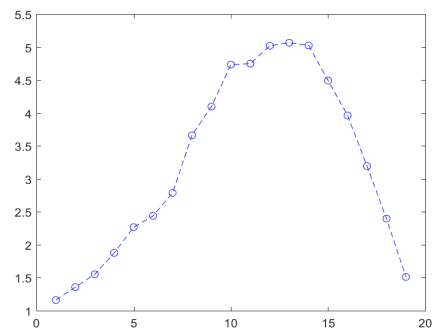
are accordingly to be segmented out. While we have earlier seen that for the smaller grain sizes or those captured at lower resolution (Fig. 4.11a), the performance indices (Table 4.2), opine in favor of vision-in-strong-contrast mode as the appropriate filter (P), for these larger grains or those captured at higher resolution here (Fig. 4.12a), the appropriate filter is generally PM, i.e. the vision-with scrutiny mode as evident for this example from the metric scores provided in Table 4.3. We again note that the results are much better as compared to those in Table 3.2 in chapter 3. However, these are the situations where the manual segmentation is less tedious because of the prominent and easily identifiable larger grains (images like Fig. 4.12a). It can also be seen from Table 4.3 that filtering is less imperative here for automated segmentation too, as compared to images like Fig. 4.11a consisting of the numerous and less distinguishable smaller grains.



(a)

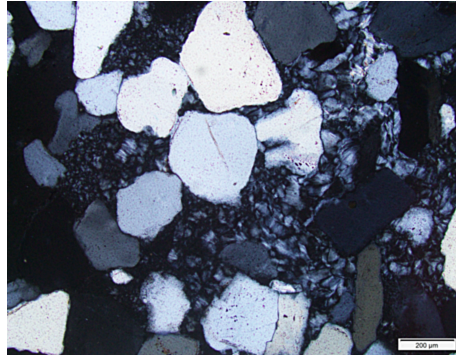


(b)

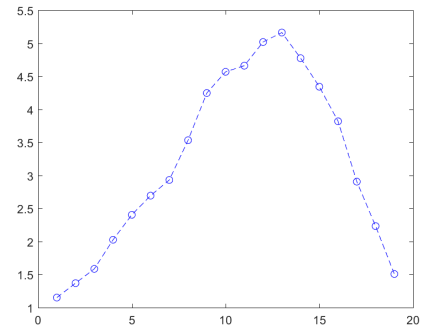


(c)

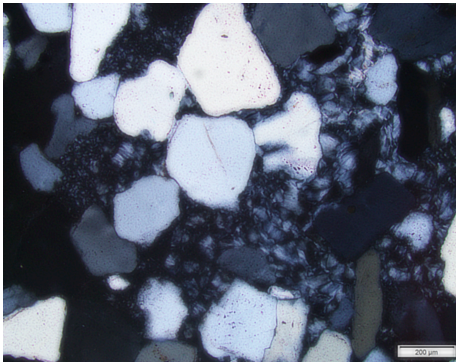




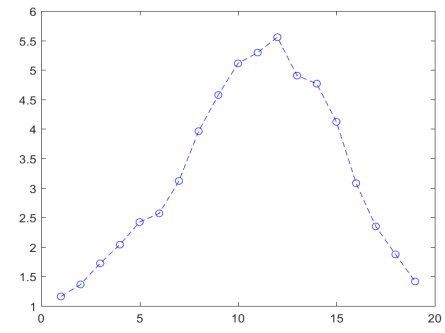
(d)



(e)

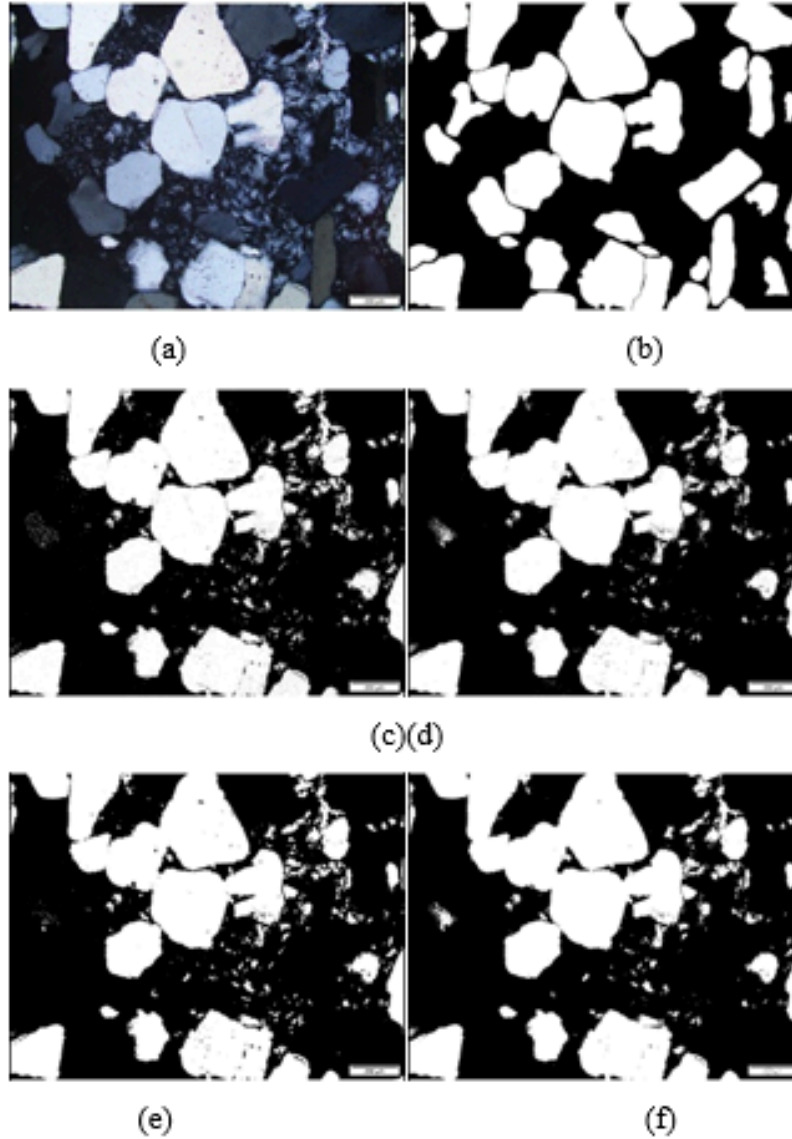


(f)



(g)

**Fig 4.12:** Results on another typical cross-polar photomicrograph (Fig 4.12a) with larger grains in the image that has been captured at higher resolution displaying its VG, VSC, and VS modes (Fig 4.12b, 4.12d, 4.12f) and their corresponding plots of cluster number (after sorting the clusters in ascending order of their centroid norms) vs.  $\beta$  index plot (Fig. 4.12c, 4.12e, 4.12g), for binary segmentation of the image after clustering into 20 clusters and optimizing for 20 runs.



**Fig 4.13:** (a) The same cross-polar photomicrograph (Fig 4.12a) displaying larger grains. (b) Ground truth of binary segmentation marked by experts. (c) The binary segmented version (optimum in 20 runs) by the mentioned segmentation algorithm on the original image in absence of any of the filters. (d) The binary segmented version (optimum in 20 runs) by segmentation algorithm on the M filtered image. (e) The segmented version (optimum in 20 runs) by the mentioned segmentation algorithm of the P filtered image (f) Binary segmented version (optimum in 20 runs) by proposed algorithm of the PM filtered image.

| Vision mode/ Other methods                            | Dice score     | Accuracy       | Jaccard Index  |
|-------------------------------------------------------|----------------|----------------|----------------|
| Without filter with modified K-means++                | 0.72976        | 0.78368        | 0.57451        |
| At a glance (M filter)<br>with modified K-means++     | 0.73925        | <b>0.79313</b> | 0.58635        |
| Strong contrast (P filter)<br>with modified K-means++ | 0.73322        | 0.78720        | 0.57881        |
| With scrutiny (PM filter)<br>with modified K-means++  | <b>0.73981</b> | 0.78877        | <b>0.58706</b> |
| Without filter with K-means                           | 0.73021        | 0.78482        | 0.57507        |
| At a glance (M filter)<br>With K-means                | 0.73937        | <b>0.79396</b> | 0.58651        |
| Strong contrast (P filter)<br>With K-means            | 0.73372        | 0.78638        | 0.57943        |
| With scrutiny (PM filter)<br>With K-means             | <b>0.74002</b> | 0.78897        | <b>0.58732</b> |

**Table 4.3:** Segmentation indices corresponding to Fig 4.13a for the M, P, and PM filters and comparison with other methods.

**Analysis:** It may be noted from Tables 4.2 and 4.3 that the proposed bio-inspired filters with K-means++ perform reasonably well. This is demonstrated using three metrics: Dice-score, Accuracy and Jaccard Index. The figures in bold font represent the highest metric values corresponding to the various filters and methods compared in the two tables, whereas those in bold-italic font represent the second highest values of the metrics. However, for smaller grains the performance is consistently good for the P filter with K-means++ as expected, while for the larger grains, results with K-means are also comparable. This is expected in the normal manual segmentation also, where it is easier to segregate the larger grains than the smaller ones. Since the K-means++ is an enhanced version, it is better to apply it in general instead of K-means to overcome some of the drawbacks of the K-means mentioned earlier. We may compare the results with the existing methods in computer vision as already performed in chapter 3. We find



that the psychophysics based models outperform the classical machine learning models. The detailed results for all the cross-polar images incorporating the different performance measures is given in Appendix A.

## 4.4 Conclusion

Grain segmentation of petrographic thin section photomicrographs is a crucial starting point for computer aided mineral identification and rock naming. These digital images contain several mineral grains and the differences among adjacent grains are usually ambiguous, especially with respect to their luminosity information, which makes current image segmentation technologies mostly inefficient. The same thing when done by an expert under a microscope, though tedious and time consuming, is quite accurate and robust. The luminosity of an object is perceived in the brain as brightness/lightness with respect to its neighborhood intensities. Although human perception based holistic models like SLIC-based methods have been applied earlier for grain segmentation, this is the first time, to our knowledge, that a brightness perception model in psychophysics has been used to achieve automatic grain segmentation with reasonably high accuracy. The results show that for images with smaller-sized grains or when captured at a lower resolution, mimicking the “vision-in-strong-contrast” mode is most effective, while for the larger-sized grain-containing photomicrographs or those captured at higher resolution, the “vision-with-scrutiny” mode is the effective strategy. We have also found that a modified version of the K-means++ clustering along with  $\beta$  index-based optimization technique for binary thresholding leads to effective automatic grain-matrix segmentation when combined with the psychophysics inspired filtered outputs. The methodology has the promise to get extended in two possible directions of an adaptive combination of these vision modes, VG, VS, and VSC as proposed in this work, developing enhanced low-level or mid-level neural models on one hand (which is the work done in the next Chapter, i.e. Chapter 5), as also incorporating high-level and holistic deep learning techniques, on the other (which is the work done in Chapter 6). In the current work, we encountered the grain-matrix segmentation from a psychophysically inspired computational model and showed that variability of photo capturing resolution or grain sizes in cross polar images plays a crucial role in proper action in our visual system to visualize that subjective cross-polar image. Unlike the current chapter which is motivated by low or early

mid-level HVS, in the next two chapters we shall investigate the same problem by applying similar-to mid-level (following David Marr's approach) and higher level vision in HVS inspired deep learning based algorithms and/or models. Although the method is found to be quite successful when applied to our dataset, there are cases of exception where the pre-processing model is not exactly followed.

# Chapter 5

## The Mid-level Vision Model of Adaptive Scrutiny for Automated Grain Segmentation in cross-polarized photomicrographs

### **Publication:**

*Das, R., Ghosh, K., 2022. Solving a difficult figure-ground segmentation problem in sedimentary rock photomicrographs using a fully automated algorithm combining Marr's Raw Primal Sketch and a Magno-Parvo Additive model. Proceedings of the 44th European Conference on Visual Perception-ECVP 2022, Nijmegen, Netherlands, August 28-September 1, 2022 (Perception, Volume 51, Issue 1\_suppl, <https://doi.org/10.1177/03010066221141167>).*

### 5.1 Introduction

In the previous chapter we attempted to segment the grains from the matrix by mimicking the eye of the geologist behind the petrographic microscope. For this we proposed that there exist two basic modes of vision, viz. Vision with Strong Contrast (VSC), and Vision at a Glance (VG), and we observed that while dealing with photomicrographs containing smaller sized grains, the VSC mode is very effective, whereas for those containing the larger grains, spatial filtering in VSC mode on a VG filtered image is more effective, which we termed as Vision with Scrutiny (VS) in the previous chapter.

We also saw in the previous Chapter that the accuracy of the proposed HVS inspired Computer Vision algorithm is higher as compared to several unsupervised machine learning methods. While the average Accuracy measured for our proposed method in chapter 4 is found to be about 77% (for the appropriate visual channel), and at least about 74% even if the appropriate channel

is missed, those for the machine learning methods in chapter 3 are lower (about 65% neglecting the much poorly performing LBP method). In this chapter, before we discuss our proposed modification of the model presented in the previous Chapter towards a more automatic segmentation, we first further verify that some other well-known and successful segmentation techniques from digital image processing literature also fail to accurately segment the grains from the matrix using our dataset. The compared methods are (i) [Kapoor et al., 1985], (ii) [Li and Lee, 1993], (iii) [Rosenfeld and Torre, 1983], (iv) [Yen et al., 1995], (v) [Abutaleb, 1989]. A comparison table is provided below.

| Thresholding methods     | Average Quantitative Scores |               |               |               |               |
|--------------------------|-----------------------------|---------------|---------------|---------------|---------------|
|                          | Dice                        | Accuracy      | Recall        | Precision     | Jaccard Index |
| Kapur et al. 1985        | 0.6316                      | 0.6391        | 0.5062        | 0.9009        | 0.4688        |
| Yen et al. 1995          | 0.6703                      | 0.6620        | 0.5661        | 0.8808        | 0.5130        |
| Rosenfeld and Torre 1983 | <b>0.7725</b>               | <b>0.7213</b> | <b>0.7670</b> | 0.8115        | <b>0.6353</b> |
| Li and Lee 1993          | 0.6667                      | 0.6685        | 0.5321        | 0.9116        | 0.5042        |
| Abutaleb 1989            | 0.6414                      | 0.6476        | 0.5061        | <b>0.9155</b> | 0.4759        |

**Table 5.1** Average performance of some of the well-known existing segmentation techniques for 16 cross-polar images other than those already compared to the proposed method in Chapter 4, in terms of dice score, accuracy, recall, precision, and jaccard index. The figures in bold indicate the best results for each of these metrics.

From Table 5.1, we find that the accuracy of all these methods fail to significantly improve upon the grain segmentation accuracy of our proposed HVS inspired model. The figures in bold indicate the best results for each of the metrics.

This motivates us in this chapter to make an attempt to make the process of mixing of channels more automatic and adaptive through a mid-level vision model, and term the resultant filter or visual mode of the expert geologist as Vision with Adaptive Scrutiny (VAS). While the low-level vision is responsible for extracting features like edges, lines, blobs etc., and the high-level for object recognition, the intermediate mid-level plays a crucial role in semantic segmentation,

according to David Marr's [Marr, 1982] representational hierarchy of visual information in the brain.

## **5.2 The Raw Primal Sketch of David Marr**

Possibly the most important contribution of David Marr, who may be called the father of neuro-visual algorithms in computer vision, was the proposal of a representational framework for vision and about deriving shape information from the two-dimensional image. Marr's low-level raw primal sketch operator [Marr, 1982], popularly called Marr-Hildreth operator [Marr and Hildreth, 1980], also known as the Laplacian of Gaussian operator forms the plank of the adaptive model of grain segmentation that we propose here.

[Marr and Hildreth, 1980] later looked upon the visual system as a complex information processing system that starts with a raw primal sketch of the external world. They considered the Laplacian, which is an orientation-independent second order differential operator and showed an equivalence of the visual receptive field, the Difference of Gaussian (DoG), discussed about in the previous chapter to the Laplacian of Gaussian (LoG) for a certain ratio of the scale parameters (i.e., the standard deviations) of the two Gaussians. This question of scale is very important for this class of filters, and shall be discussed later in this chapter in the context of the present model. According to the Marr-Hildreth scheme, the zero-crossings of the image filtered with LoG are likely to be detected at the simple cell level of the primary visual cortex, and form the raw primal sketch of the external world.

## **5.3 The Proposed Algorithm for Adaptive Scrutiny**

In our proposed mid-level vision model, we first apply the M filter to get the Vision at a Glance (VG) image, following the same procedure for construction of the PM filter in the previous chapter. But here, instead of directly applying the P filter on it, we follow the scheme as proposed by [Marr, 1982]. We use the Laplacian of Gaussian operator at every pixel of the VG

image and calculate the corresponding squared (positive) values. In other words, we calculate the edge strengths at each position of the image. If we divide the total edge strength of the image by the total number of pixels, we get the average edge strength of the image per pixel. Each image thus gets characterized by a number corresponding to its raw primal sketch. Using a function  $f(\mu) = \mu / (\mu + k)$ , where  $k$  is a constant (positive integer), we map the values of  $\mu$  to  $[0, 1)$  interval. So for each photomicrograph corresponding to its raw primal sketch, we get a value of  $f(\mu) = m_h$  (a variable signifying the Marr-Hildreth operator )

We now linearly combine the VG and VSC using the  $m_h$  variable as:

$$VAS = m_h * VG + [1 - m_h] * VSC$$

This enables us to arrive at the desired mid-level vision model, where the Vision with Scrutiny mode defined in the previous Chapter becomes the Vision with Adaptive scrutiny (VAS). We describe the step-by-step overall mid-level vision algorithm below:

**Algorithm 5.1 : mid-level vision algorithm**

- Step 1: Read input Image I.
- Step 2:  $VG = M\_filter(I)$ . (Apply M-filter on I to obtain the Vision at a Glance (VG))
- Step 3:  $L = \nabla^2(VG)$ . (At each pixel of VG, we calculate the Laplacian of Gaussian to determine level of detail in the raw primal sketch)
- Step 4:  $\mu = Mean(L^2)$ . (Compute mean square value of L per pixel.  $\mu$  is a measure of the detail in the raw primal sketch)
- Step 5: Compute  $f(\mu) = \mu / (\mu + k)$ . (Maps  $\mu \in [0, \infty)$  to  $f(\mu) \in [0, 1)$  )
- Step 6:  $VSC = P\_filter(I)$ . (Apply P-filter on I)
- Step 7:  $VAS = m_h * VG + [1 - m_h] * VSC$  ( $m_h$ ) (where  $f(\mu) = m_h$ , the Marr-Hildreth variable)

We thus get a mid-level vision model Adaptive Scrutiny based on the low-level extraction of raw primal sketch using the Marr-Hildreth operator, by connecting the two primary visual modes in the two parallel visual channels through the Marr-Hildreth variable  $m_h$  that we have defined

here. The model clearly demonstrates that more detailed raw primal sketch implies higher value of  $m_h$ , and hence less dependence on P, i.e. on the strong contrast mode of vision when performing the Adaptive scrutiny through the two primary visual modes, i.e. VG and VSC.

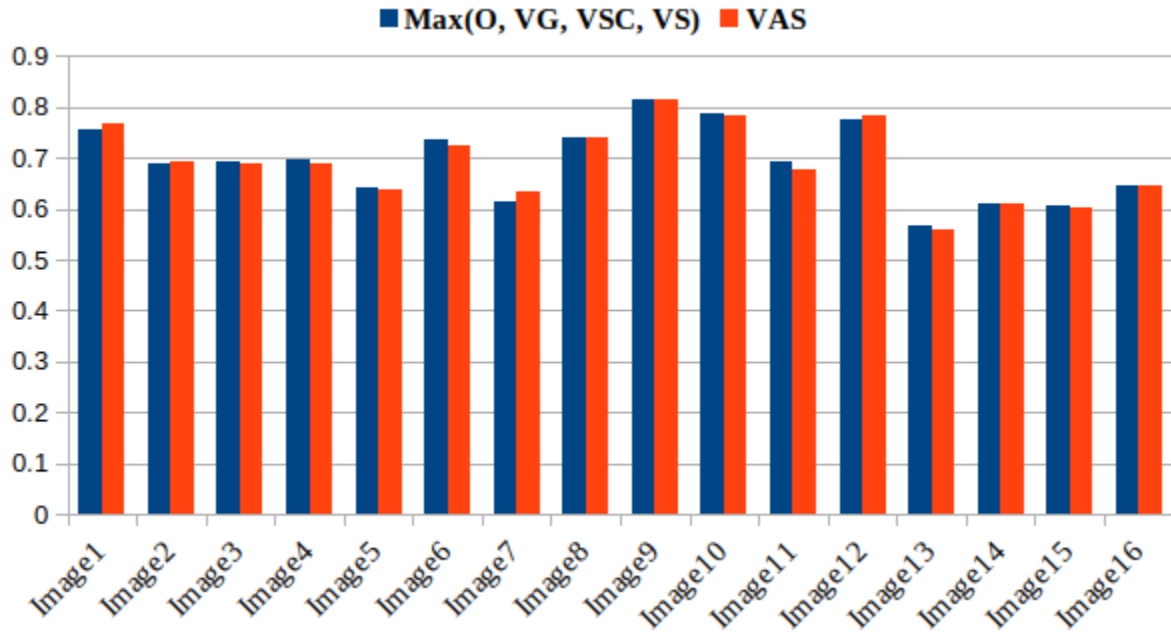
## 5.4 Results

In spite of some variation over scale, the LOG based model of the current method appears to be a potential candidate for automating channel selection. This is clearly visible from Table 5.2. It can be seen that the variation of Dice scores over variance of the LOG, is not very significant, and

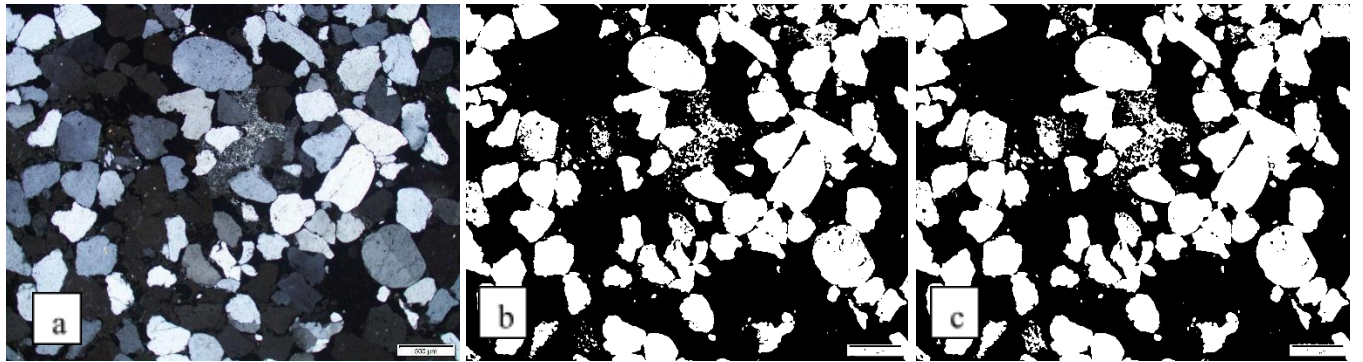
|          | $\sigma^2 = 2$ | $\sigma^2 = 4$ | $\sigma^2 = 6$ | $\sigma^2 = 8$ | $\sigma^2 = 10$ | $\sigma^2 = 12$ | $\sigma^2 = 14$ | $\sigma^2 = 16$ |
|----------|----------------|----------------|----------------|----------------|-----------------|-----------------|-----------------|-----------------|
| Image 1  | 0.7527         | 0.7648         | 0.7657         | 0.7672         | 0.7674          | 0.7674          | 0.7674          | 0.7674          |
| Image 2  | 0.6691         | 0.677          | 0.6923         | 0.6897         | 0.6897          | 0.6898          | 0.6898          | 0.6898          |
| Image 3  | 0.6771         | 0.6883         | 0.689          | 0.689          | 0.689           | 0.689           | 0.689           | 0.689           |
| Image 4  | 0.6884         | 0.6253         | 0.6326         | 0.6322         | 0.6322          | 0.6322          | 0.6322          | 0.6322          |
| Image 5  | 0.6165         | 0.6228         | 0.6382         | 0.6382         | 0.6382          | 0.6382          | 0.6382          | 0.6382          |
| Image 6  | 0.7234         | 0.7132         | 0.7164         | 0.7201         | 0.7165          | 0.7165          | 0.7165          | 0.7165          |
| Image 7  | 0.5904         | 0.6339         | 0.5904         | 0.5903         | 0.5903          | 0.5903          | 0.5903          | 0.5903          |
| Image 8  | 0.7381         | 0.7356         | 0.7337         | 0.7339         | 0.7339          | 0.7339          | 0.7339          | 0.7339          |
| Image 9  | 0.8159         | 0.8157         | 0.8152         | 0.815          | 0.815           | 0.815           | 0.815           | 0.815           |
| Image 10 | 0.7846         | 0.7802         | 0.7798         | 0.7796         | 0.7799          | 0.7799          | 0.7799          | 0.7799          |
| Image 11 | 0.6735         | 0.6771         | 0.6754         | 0.6733         | 0.6733          | 0.6733          | 0.6733          | 0.6733          |
| Image 12 | 0.7736         | 0.7815         | 0.7777         | 0.7767         | 0.7767          | 0.7766          | 0.7766          | 0.7766          |
| Image 13 | 0.5595         | 0.4983         | 0.4820         | 0.4823         | 0.4817          | 0.4817          | 0.4817          | 0.4817          |
| Image 14 | 0.6085         | 0.6091         | 0.6064         | 0.6069         | 0.6057          | 0.6067          | 0.6069          | 0.6069          |
| Image 15 | 0.6012         | 0.6023         | 0.6021         | 0.6021         | 0.6021          | 0.6021          | 0.6021          | 0.6021          |
| Image 16 | 0.6401         | 0.6412         | 0.6386         | 0.6437         | 0.6400          | 0.6400          | 0.6400          | 0.6400          |

**Table 5.2:** Table showing the Dice score results for our dataset with Gaussian variance varying from 2 - 16. It is found that the variation of Dice scores over variance of the LOG, is not very significant, and most of the results are near optimal if we choose variance = 4 for our automated model.

most of the results are near optimal if we choose variance = 4 for our automated model.



**Fig. 5.1:** Comparison of the automated channel selection method VAS with those in the previous Chapter (Original, VG, VSC, and VS) in terms of Dice scores.



**Fig. 5.2:** a. Original cross polar image, b. Segmented image by non-automatic visual channel method, c. Segmented image by the currently proposed LOG-based automatic method

In Fig. 5.1, we compare the best results from the non-automatic channel selection in the previous Chapter with those of the LOG-based automatic method (across variance) in terms of Dice scores. Both Table 5.1 and Fig. 5.1 demonstrates that the motivation for automating the visual channel based method stands justified.



In Fig. 5.2, we show results of segmentation in a sample XPL image using the non-automatic optimal VSC filter and the LOG-based VAS method at variance=4 of the Gaussian in LOG.

## 5.5 Conclusion

In this Chapter, we develop a new model of mid-level vision, namely Vision with Adaptive Scrutiny using which automate the task of choosing the appropriate visual channel based filter described in the previous chapter, for pre-processing the sedimentary rock photomicrographs before applying the clustering based segmentation algorithm. It basically applies the raw primal sketch operator LOG on the Vision at a Glance (VG) image, and calculates a parameter which is used to linearly combine the VG and VSC modes. It is seen that such an adaptive filter may be used with reasonably good accuracy in a fully automated mode, across various variances, but if we choose variance = 4 of the Gaussian in the LOG operator, then it is almost optimal across our dataset. This automated method thus improves upon, in general, the semi-automated version discussed in the previous chapter. In Appendix B, we have presented the confusion matrices for pixel classification for the different unsupervised methods discussed so far on a typical photomicrograph from our dataset.

In the next Chapter, we now move on to the state of the art in supervised segmentation, that is application of Deep Neural Networks to train and segment the photomicrographs into grain and matrix regions. It should be kept in mind that this data-driven deep learning architecture is also inspired by the hierarchy in the mammalian and primate visual cortex. The deep learning models that we mostly use in the next chapter consist of an encoder and decoder architecture, which also has many similarities with the high resolution VSC channel, and low resolution VG channel in HVS, we are trying to model in this thesis.

# Chapter 6

## Deep Learning Methodology for Automatic Grain-matrix Segmentation in Combined Plane and Cross-polarized Sandstone Photomicrographs

### ***Publications:***

*Das, R., Mondal, A., Chakraborty, T., & Ghosh, K. (2022). Deep neural networks for automatic grain-matrix segmentation in plane and cross-polarized sandstone photomicrographs. Applied Intelligence, 52(3), 2332-2345.*

*Das, R., Chowdhury, A., Chakraborty, T., & Ghosh, K. (Accepted for publication in SPIE Proceedings). SPGMS: An Annotated Benchmark for Grain-Matrix Segmentation in Sandstone Photomicrographs. ICVIP 2024 (The 8th International Conference on Video and Image Processing) which will be held on December 13-15, 2024 in Kuala Lumpur, Malaysia.*

### **6.1 Introduction**

In Chapter 3 we introduced a novel dataset for grain matrix segmentation and tested it with various unsupervised machine learning models. These classical machine learning models are used in various computer vision tasks. However, we noted in Chapter 3 that these models did not achieve the desired results as we had also predicted in our introductory chapters in this thesis citing the level of difficulty in the grain matrix segmentation problem. Hence in chapters 4 and 5, we tried to provide novel psychophysics based models for processing the XPL images of our dataset in line with David Marr's algorithmic approach.

This chapter delineates our research on synthesizing new images from plane polar (PL) and cross polar (XPL) images through their amalgamation, as well as the application of deep learning techniques for image segmentation in sandstone photomicrographs. Image segmentation, as has already been discussed in Chapter 2, is a basic and crucial step in many image analysis and computer vision tasks. The task of image segmentation sometimes reduces to the form of binary classification, which is often referred to as figure-ground or object-background segmentation, as is the problem of grain-matrix in this thesis. Object-background segmentation has many important applications like document analysis [He et al., 2019], [Vincent et al., 2018], visual searching [Wolfe et al., 2002], saliency detection [Hou and Zhang, 2007], and video tracking [Chan, 2018]. In the segmentation of the objects from the background, the goal is to divide the image into two homogeneous partitions based on luminance, color, texture, etc. Choosing the appropriate threshold in image histograms is often an effective method for such object-background partitioning. However, this simple binary classification task can turn out to be intriguing, especially in situations involving uncertainties and ambiguities with color, luminance, and/or texture information, or when multiple objects are to be detected from their background and such objects have overlapping boundaries. In this chapter, we attempt to test our novel dataset with several state of the art deep learning segmentation techniques to overcome these difficulties.

Deep learning has transformed computer vision in general and image segmentation in particular by offering automated, high-precision algorithms that surpass conventional computer vision techniques. The increasing use of deep learning methodologies for segmentation is mainly due to the efficacy of convolutional neural networks (CNNs) at acquiring spatial hierarchies and intricate characteristics from data. In this chapter, we apply various deep-learning methods on our existing datasets and compare them. Apart from simply applying the models we introduce a new loss function called Weighted Binary Cross Entropy Loss function and show that replacing this loss function with simple Binary Cross Entropy or Dice-Score loss function results in improved performance.

The segmentation of grains in a sandstone thin-section under the plane and cross-polarized light in a microscope is a fundamental step in this study. The grains are the larger fragments to be segregated from the matrix (the finer-grained sedimentary material, such as clay, silt or

cementing material that surrounds or occurs in the interstices of larger grains) of the rocks. The geologist usually identifies the grains by studying a thin section observed under a polarizing microscope or a photomicrograph image of the thin section.

A photomicrograph typically contains hundreds of mineral grains in which the manual segmentation is very tedious, time-consuming, and subjective to individual expertise. This is especially so, in view of the overlapping grain boundaries in the two-dimensional images, as well as the large scale variation of color and intensity of the grains/clasts depending upon several factors like the crystal structure of the minerals, orientation of the crystal axes with respect to the plane of microscope stage, thickness of mineral grains and other optical properties of the grain. In plane polarized light in a petrographic microscope (only with the basal polariser used) one can see certain properties of the grains like refractive index, body color etc., whereas in a cross-polar view (with both the polariser and the analyser inserted), one can observe properties like birefringence, extinction position etc. In the case of isotropic materials (like glass, water, mineral garnet), under cross polar view, all the light transmitted through the mineral is blocked by the analyser and the grain appears dark in all positions of the rotating microscope stage. However, in case of anisotropic minerals, in a cross-polar view one can observe light of different brightness and color in different grains of the same mineral (due to variation in the orientation of the crystals in different grains). As the microscope stage is rotated, the brightness changes continuously. Thus an anisotropic mineral grain (quartz, feldspar etc) in one position may show bright gray color while in the other position it may be completely dark. Further, due to various other processes operating during the transformation of loose sand to sandstone, the boundaries of grains can be modified to be extremely complex and often fuzzy, hindering easy delineation of the grain boundary from the surrounding matrix. This makes the task of automated grain segmentation even more challenging. The photomicrographs meant for such automated segmentation, depending upon the orientation and type of the various crystals present in the slides, display wide variance in grain luminance from almost invisibly dark to very bright in appearance. Yet, such apparent variation may in fact often be misleading, a fact which makes automated grain segmentation in sandstone photomicrographs an all the more difficult task. This is why, in the absence of reliable computer vision algorithms for grain segmentation (despite several important attempts mentioned in the following section), the geologists have to combine all the features observable in plane and cross-polarized view and in different positions of the

rotating microscope stage, and have to perform the segmentation task manually and tediously. However, the earlier chapters in this thesis have only considered the more informative cross-polarized view (and not the plane-polarized images) to achieve grain-matrix segmentation either through standard machine learning algorithms or through the novel psychophysics inspired models developed in this thesis. For the first time in this thesis, we shall incorporate the plane-polarized view also in our dataset for training the deep learning models used in this chapter. This methodology of combining the plane polarized with the cross polarized view is in fact a salient contribution of this chapter so far not done by any other researcher to the best of our knowledge for achieving the automated grain matrix segmentation.

Thus, we summarize the contribution of this work as follows.

- We produce new images by averaging the information between each pair of cross polar and plane polar images.
- We introduce a new loss function called weighted binary cross entropy and incorporate it with various deep learning models and compare them with various evaluation metrics.

## **6.2 Image Semantic Segmentation using Deep Learning**

Pixel-wise classification task like image semantic segmentation achieves significant improvement inspired by replacing the fully connected layer in the image classification network with the convolution layer, referred to as Fully Convolutional Network (FCN) [Lafferty et al., 2001]. Limitations of the FCN model [Lafferty et al., 2001] include ignoring small objects and mislabeling the large objects due to fixed receptive field size. To address this issue, Noh et al., [Mingireanov Filho et al., 2013] proposed a semantic segmentation algorithm by learning a deconvolution network composed of both deconvolution and un-pooling layers. Several methods [Chen et al., 2018], [Zhao et al., 2017] used dilated convolution to enlarge the receptive field of the neural network. Since in deep networks, higher-layer features contain more semantic meaning and less location information, various methods [Chen et al., 2016], [45] have been developed by combining multi-scale features to improve the segmentation performance. Another direction in semantic segmentation is based on structure prediction. The pioneer work [Chen et

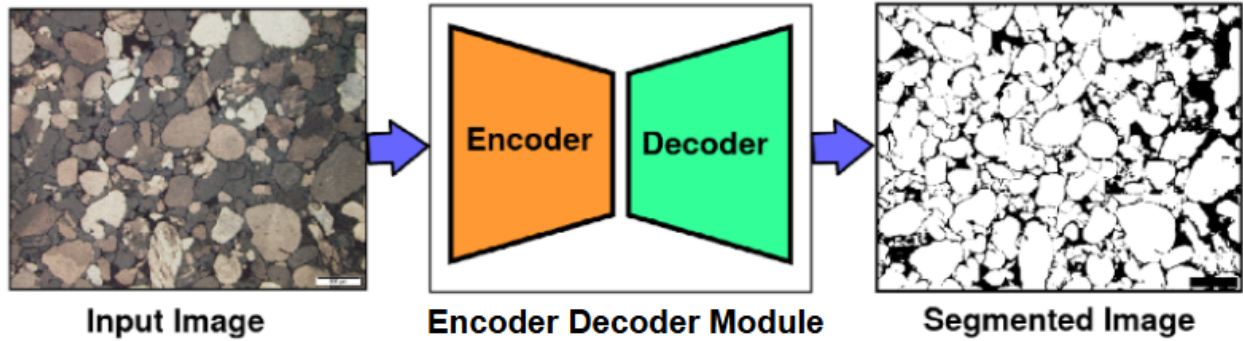
al., 2018] considers the conditional random field (CRF) [Jungmann 2014] for post-processing to refine the segmentation result. Various methods [Arnab et al., 2016], [Zheng et al., 2015] included CRF to refine the networks' end-to-end modeling. The inclusion of CRF improves the segmentation boundaries.

All these above discussed networks consist of encoders and their corresponding decoders. Some spatial information is lost in decoders because of performing multiple down-sampling operations (max-pooling) in the encoders. It is very difficult to recover this lost information in decoders using only the down-sampled outputs of encoders. Badrinarayanan et al., [Badrinarayanan et al., 2016] solved this problem by linking encoder with decoder through pooling indices which are not trainable parameters. In this direction, various architectures such as U-net [Romera et al., 2017] and U-net++ [Zhou et al., 2018] with trainable parameters have been proposed to solve this problem. However, all these discussed networks use VGG16 (138 million parameters) or RESnet101 (45 million parameters) as their encoder which are huge in terms of parameters and GFLOPS, excepting U-net++. These architectures cannot be used for real-time applications. Various networks such as ERFnet [Noh et al., 2015] and Linknet [Chaurasia 2017] are developed for real-time applications.

## **6.3 Deep Learning Frameworks For Grain Segmentation**

While visual recognition has been an active research topic for the last few decades, there has been a recent trend shift in its research from traditional hand-crafted feature design, as discussed in the previous section, to feature extraction through deep networks. Through several instances in computer vision literature, it has recently been proven that the later data-driven approach often far outperforms the hand-crafted feature-based earlier approaches. We pose grain segmentation as a pixel-wise two-class (i.e., grain and background) semantic segmentation task in this work. For this purpose, we have applied several deep learning models to achieve this segmentation task. All of these deep learning models consist of an encoder network and a corresponding decoder network followed by a pixel-wise classification layer. In general a deep learning model for semantic segmentation task takes an original image  $x$  as input, and produces a segmented

image  $y$  as output. Figure 6.1 illustrates the overview of a deep learning model (LinkNet [Chaurasia et al., 2017]).

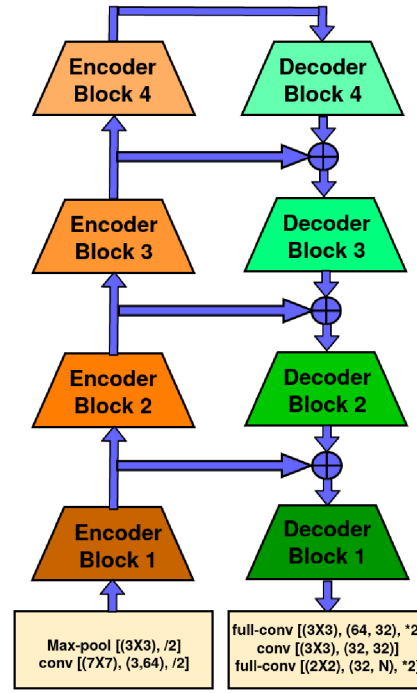


**Fig. 6.1:** Visually illustrates an overview of a deep learning model for semantic segmentation tasks. It takes an input image and produces a segmented image as output. The network's encoder learns the features while the decoder reconstructs the foreground image by up-sampling the learned features for classification.

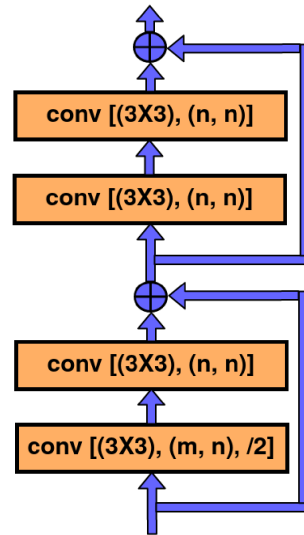
### 6.3.1 Network Architecture

Figure 6.2 displays our encoder-decoder architecture. In this figure, ‘Conv’ means convolution, ‘Full-conv’ means full convolution, and ‘/2’ indicates down-sampling by a factor 2, which is performed by stride convolution. Similarly, ‘\*2’ represents up-sampling by a factor of 2. We consider batch normalization between the convolutional layer and ReLU. The encoder-decoder network in Figure 6.2 indicates the left part as an encoder and the right part as a decoder. The encoder and the decoder, in turn, consist of several encoder and decoder blocks, respectively, as shown in Figure 6.2. The initial encoder block performs convolution on the input image with 7 x 7 kernel and a stride 2. After that, max-pooling is performed on a 3 x 3 area with stride 2. All other encoder blocks consist of a residual block [He et al., 2016], which is represented as encoder-block (i). Individual layers within these blocks are shown in Figure 6.3. Similarly, the detailed layers in the decoder-blocks are provided in Figure 6.4. We consider ResNet18 (11.5 million parameters) [He et al., 2016], a lighter network as compared to vgg-16 and ResNet101, as an encoder. Fine-tuning this network for a specific task with the limited training samples is not a problem. On the other hand, the information lost in the encoder is retained by bypassing the output of each encoder to its corresponding decoder, which is shown in Figure 6.2. In this

architecture, the decoder is sharing the knowledge learned by the encoder at every layer. The decoder uses fewer parameters resulting in an efficient network.

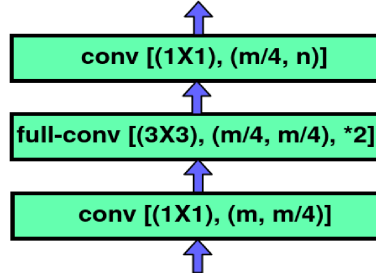


**Fig. 6.2:** Shows the encoder-decoder architecture in the Lin[LinkNet Module → Encoder-decoder module]kNet [Garcia et al., 2020] module. Left: shows the convolution modules in the encoder network, and Right: shows the convolution modules in the decoder network.



**Fig. 6.3:** Shows the details of convolution modules in the encoder-block(i).





**Fig. 6.4:** Shows the details of convolution modules in the decoder-block (i).

### 6.3.2 Loss function comparison

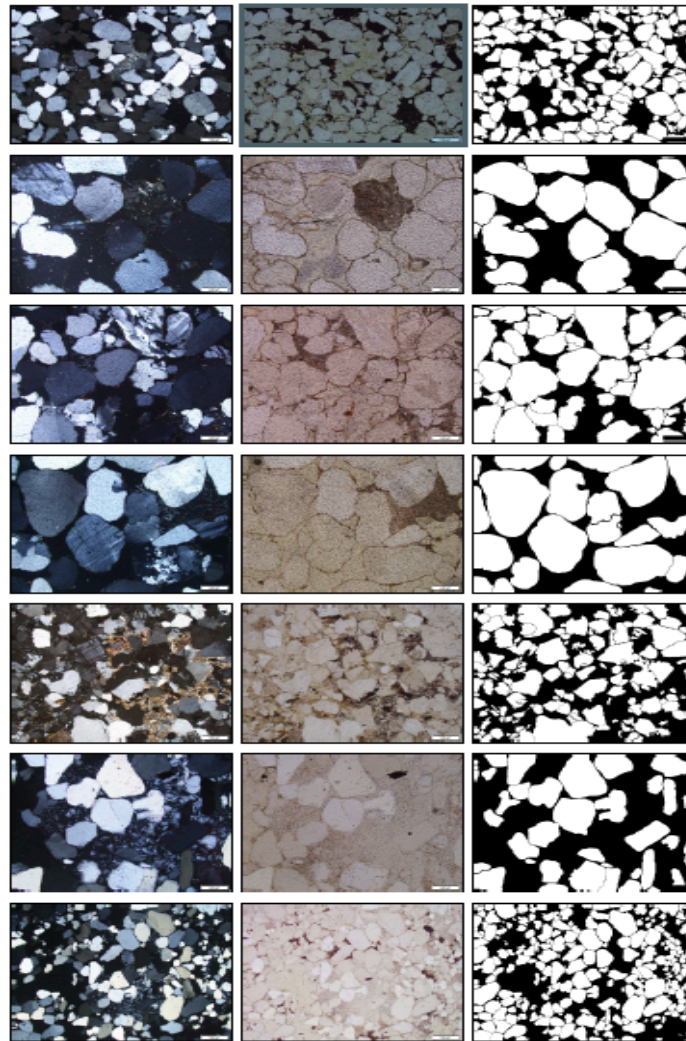
Let,  $x_{ij}$  represent the information of the input image  $x$  at pixel location  $(i, j)$ , while  $y_{ij}$  represents the class label (i.e. 1 for grain and 0 for background) at pixel position  $(i, j)$  of the output segmented image  $y$ . Let the ground truth segmented image be  $y^{gt}$  corresponding to the input image  $x$ . Therefore,  $y_{ij}^{gt}$  denotes the class label (i.e., 1 for grain and 0 for background) at pixel location  $(i, j)$  of the ground truth segmented image  $y_{gt}$ . We formulate the semantic segmentation task as  $y = f(x)$ , where  $x$  is the input image and  $y$  is the segmented image and  $f$  is the pixel-wise non-linear mapping function. During training, given a pair of images  $(x, y^{gt})$ , the network learns this mapping function to the network's weight parameters. The mapping function  $f$  is represented with the network's weight parameters learned by minimizing pixel-wise weighted two-class cross entropy loss between the predicted segmented image  $y$  and ground truth segmented image  $y^{gt}$ . We have applied our datasets (discussed in Chapter 3) to several such deep learning models to train and compared the results. Apart from the loss function such as Binary Cross Entropy Loss (BCEL) or Dice Score Loss (DSL), we have introduced a new loss function Weighted Binary Cross Entropy Loss function (WBCEL). In table 6.1 we have defined these loss functions.

| Loss Function                              | Formula                                                                                                                                                                                                                    |
|--------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Binary Cross Entropy Loss (BCEL)           | $E = \sum_i \sum_j -y_{ij} \log(y_{ij}^{gt}) - (1 - y_{ij}) \log(1 - y_{ij}^{gt})$                                                                                                                                         |
| Dice Score Loss (DSL)                      | $1 - \frac{\sum_i \sum_j y_{ij} y_{ij}^{gt} + \epsilon}{\sum_i \sum_j y_{ij} + y_{ij}^{gt} + \epsilon} - \frac{\sum_i \sum_j (1 - y_{ij})(1 - y_{ij}^{gt}) + \epsilon}{\sum_i \sum_j 2 - y_{ij} - y_{ij}^{gt} + \epsilon}$ |
| Weighted Binary Cross Entropy Loss (WBCEL) | $E = \sum_i \sum_j -w_1 y_{ij} \log(y_{ij}^{gt}) - w_0 (1 - y_{ij}) \log(1 - y_{ij}^{gt})$                                                                                                                                 |

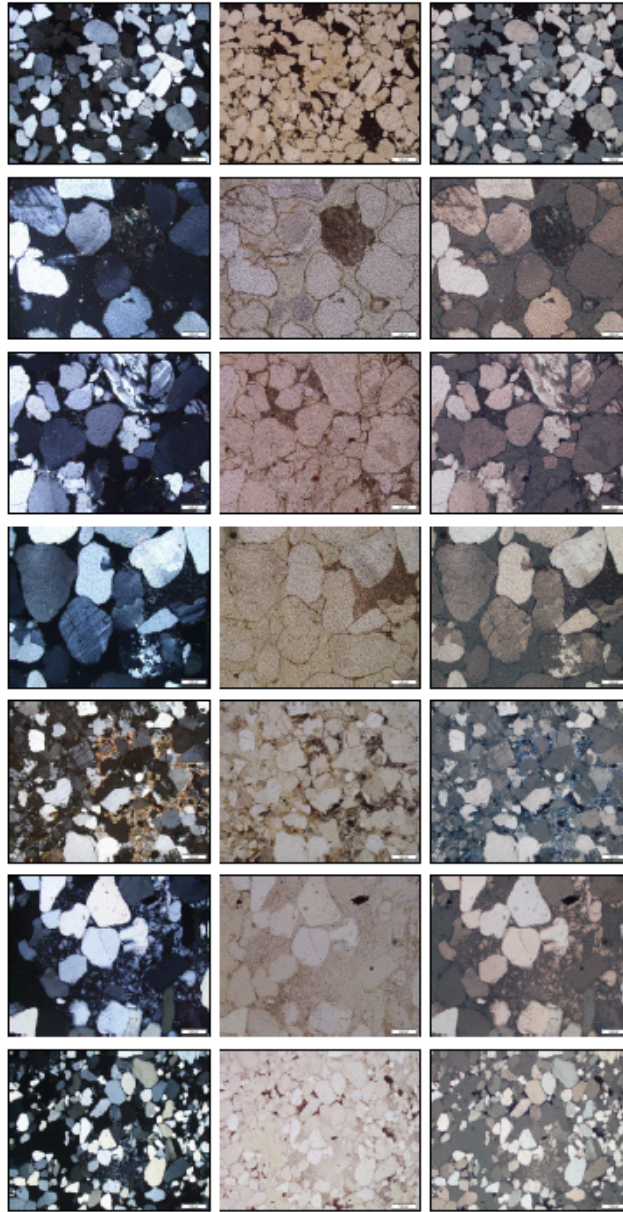
**Table 6.1:** Comparing various loss functions, where  $w_1 = \frac{Nobj}{N}$  and  $w_0 = \frac{Nbg}{N}$  with  $Nobj$ ,  $Nbg$  and  $N$  are the number of pixels within the grain region, number of pixels within the background region and the total number of pixels in an image of the training set, respectively. The term  $\epsilon$  is used to ensure the dice loss function stability by preventing the issue of dividing by 0.

## 6.4 Experiments

Out of the nineteen images in the SPGMS dataset, as already mentioned earlier, we have used sixteen triplets, consisting of the cross polar, the plane polar, and the ground truth images in these experiments (For the entire work done in this thesis, we have not used the other three images because one of them has no corresponding plane polar image, and for the other two sample images, the ground truth images are ternary containing ambiguous regions in them). Some of these are demonstrated in Figure 6.5.



**Fig. 6.5:** Illustrates some of the images from our dataset, used in this current work, with ground truth segmentation. First Column: shows images in XPL. Second Column: shows same images in PPL and Third Column: shows corresponding ground truth segmentation.



**Fig. 6.6:** Shows the generated average image from some pair of images (cross-polar and plane-polar), used in this current work, from our dataset. First Column: shows images in XPL, Second Column: shows images in PPL and Third Column: shows the generated average image of First Column and Second Column.

### 6.4.1 Implementation Details

**Pre-processing:** We have only ten pairs of images for training purposes. It is challenging to train our neural networks with these limited (i.e., pair of 10) training images. We apply various pre-processing steps to these training images to generate a larger number of images for training the network. Since the network takes an input image of size 256 x 256, we, first of all, crop 256 x 256 regions from the original images to generate a large training set. Here, we consider that each image is independent of the other for a single pair.

- **Training-Set1:** We crop non-overlapping 256 x 256 regions from the images of the training set containing a pair of each of the ten images and thus create a new training set, named as Training-Set1. The Training-Set1 consists of 1408 cropped images. Since the original image sizes are not wholly divisible by the cropped image size 256 x 256, we use padding with 0 to include the boundary regions into the Training-Set1.
- **Training-Set2:** We crop 256 x 256 regions with 128 pixels overlapping from the images of the training set to create a new training set, named as Training-Set2. It contains 5142 cropped images. We use padding with 0 to include the boundary regions into the Training-Set2.
- **Training-Set3:** Similarly, we crop 256 x 256 regions with 64 pixels overlapping from the images of the training set to create a new training set, named as Training-Set3. It contains 2348 cropped images. We use padding with 0 to include the boundary regions into the Training-Set3.
- **Test-Set1:** We crop non-overlapping 256 x 256 regions from the images of the test set containing six pairs of images, to create a new test set, named Test-Set1. We apply padding of 0 to include the boundary regions into the set. The Test-Set1 contains 768 images.

As another contribution, we create another group of training sets by considering the dependency between images of a single pair in the actual training set containing ten pairs of images. We produce ten new images by averaging the information between each pair of images. Similarly, we create a test set containing six images by averaging individual pairs of images in the original

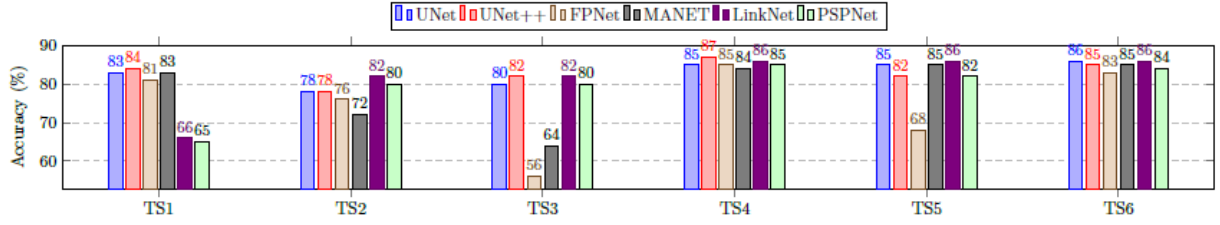
test set. Figure 6.6 shows some such average images. We then create various training sets by cropping 256 x 256 regions from the images.

- **Training-Set4:** We crop non-overlapping 256 x 256 regions from the images of the training set containing ten average images and create a new training set, named as Training-Set3. The Training-Set3 consists of 704 cropped images. We use padding with 0 to include the boundary regions into the Training-Set3.
- **Training-Set5:** We crop 256 x 256 regions with 128 pixels overlapping from the ten average images to create a new training set, named Training-Set5. It contains 2571 cropped images. We use padding with 0 to include the boundary regions into the Training-Set5.
- **Training-Set6:** Similarly, we crop 256 x 256 regions with 64 pixels overlapping from the ten average images of the training set to create a new training set, named Training-Set6. It contains 1174 cropped images. We use padding with 0 to include the boundary regions into the Training-Set6.
- **Test-Set2:** We crop non-overlapping 256 x 256 regions from the six average images of the test set to create a new test set, named Test-Set2. We apply padding of 0 to include the boundary regions into the set. The Test-Set2 contains 384 images.

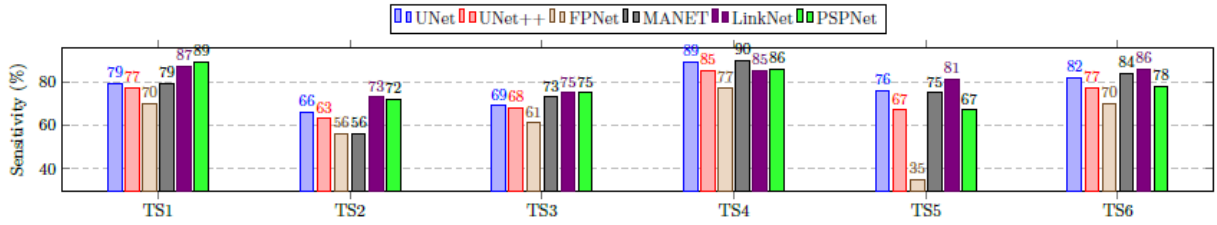
**Training details:** We considered a pre-trained ResNet-18 model (on ImageNet [Deng et al., 2009]). We do this for all the experiments. We use 3 x 3 kernels for all the convolutional layers. The stride of each pooling layer is set to 2. For the first decomposition block, we use 64 filters. The number of filters is doubled for each of the subsequent decomposition blocks. The number of filters in each deconvolutional layer is the same as its corresponding pooling layer. We train the model with the training set with batch size 16 for 60 epochs. The initial learning rate is set to 0.0005 with a decay factor of 0.1. We train our model with the various training sets with the same values of parameters.

### 6.4.2 Ablation Study on Various Training Sets

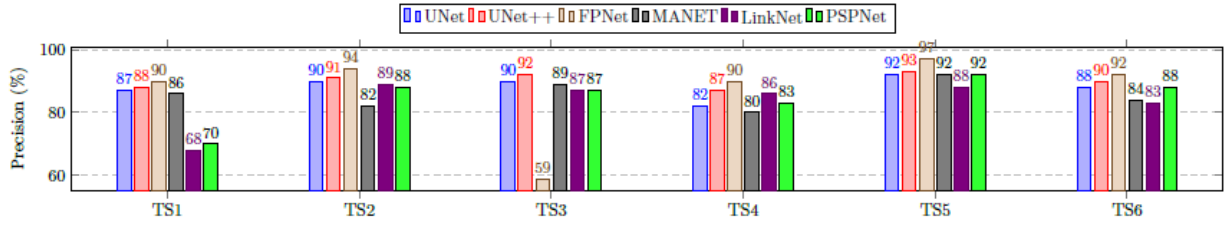
Figure 6.7 shows the performance of various deep learning models trained with different training sets while Figure 6.8 demonstrates the qualitative results obtained using the various deep learning based segmentation architectures. When a model is trained with Training set1 to 3, it is tested against Test set1. Similarly, for Training set4 to 6, the model is tested against Test set2.  $TS_i$  represents the Training set  $i \in \{1, 2, 3, 4, 5, 6\}$ . Figure 6.7 (a) represents the segmentation accuracy of various deep learning models. The scale ranges from a minimum of about 56% to a maximum of around 87%. UNet and UNet++ show a consistent performance. For both FPNet and MANET, performance drops on TS3. FPNet has suboptimal performance on TS5. LinkNet and PSPNet perform consistently except on TS1. In the case of Figure 6.7 (b), FPNet shows low recall or sensitivity on TS5 (35%). Other models show consistent performance throughout the Training datasets. Figure 6.7 (c) shows a comparison of precision among the models. UNet, UNet++, and MANET show consistent performance among all training datasets. FPNet exhibits a decline in performance on TS3 (59%). LinkNet and PSPNet show good scores among all the datasets except TS1. In the case of the Jaccard index (Intersection over union), all the models except FPNet show consistent performance across the datasets as shown in Figure 6.7.



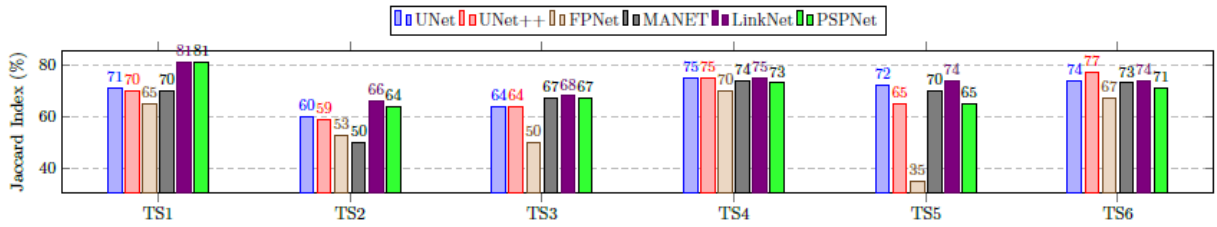
(a) Accuracy of segmentations generated by various state-of-art deep-learning methods.



(b) Sensitivity of segmentations generated by various state-of-art deep-learning methods.



(c) Precision of segmentations generated by various state-of-art deep-learning methods.



(d) Jaccard Index of segmentations generated by various state-of-art deep-learning methods.

**Fig 6.7:** Comparison of various state of the art deep learning frameworks on our dataset. TS<sub>i</sub> denotes the Training set  $i \in \{1,2,3,4,5,6\}$ .



### 6.4.3 Ablation Study on Loss Function

Table 6.2 shows the obtained quantitative results by the used architectures. The yellow part shows the result of various deep learning methods using the Binary cross entropy loss (BCEL), the blue part shows the result using Dice Score Loss (DSL), while the green part shows the result of using the weighted binary cross entropy loss (WBCEL). While comparing, the result indicates that when we plug the models with our proposed loss function, we achieve better results in most of the cases. U-Net++ obtains the best performance across all cases. LinkNet with WBCEL is able to produce results comparable with U-Net++.

| Method  | Loss Function | Quantitative Score on Test Images |              |              |              |              |
|---------|---------------|-----------------------------------|--------------|--------------|--------------|--------------|
|         |               | Accuracy↑                         | Recall↑      | Precision↑   | Dice↑        | JI↑          |
| U-Net   | BCEL          | 0.793                             | 0.809        | 0.793        | 0.796        | 0.638        |
| U-Net++ | BCEL          | 0.832                             | 0.865        | 0.833        | 0.848        | 0.710        |
| FPNet   | BCEL          | 0.722                             | 0.745        | 0.701        | 0.722        | 0.650        |
| MANet   | BCEL          | 0.701                             | 0.738        | 0.721        | 0.753        | 0.548        |
| LinkNet | BCEL          | 0.795                             | 0.833        | 0.798        | 0.815        | 0.727        |
| PSPNet  | BCEL          | 0.766                             | 0.765        | 0.756        | 0.760        | 0.695        |
| U-Net   | DSL           | 0.781                             | 0.661        | 0.901        | 0.761        | 0.602        |
| U-Net++ | DSL           | 0.861                             | 0.742        | <b>0.911</b> | 0.812        | 0.691        |
| FPNet   | DSL           | 0.762                             | 0.561        | 0.902        | 0.701        | 0.532        |
| MANet   | DSL           | 0.724                             | 0.563        | 0.823        | 0.665        | 0.501        |
| LinkNet | DSL           | 0.823                             | 0.732        | 0.892        | 0.804        | 0.662        |
| PSPNet  | DSL           | 0.802                             | 0.721        | 0.882        | 0.792        | 0.644        |
| U-Net   | WBCEL         | 0.837                             | 0.853        | 0.837        | 0.845        | 0.682        |
| U-Net++ | WBCEL         | <b>0.889</b>                      | <b>0.919</b> | 0.890        | <b>0.904</b> | <b>0.789</b> |
| FPNet   | WBCEL         | 0.832                             | 0.801        | 0.722        | 0.759        | 0.698        |
| MANet   | WBCEL         | 0.795                             | 0.745        | 0.734        | 0.739        | 0.641        |
| LinkNet | WBCEL         | 0.841                             | 0.859        | 0.841        | 0.849        | 0.739        |
| PSPNet  | WBCEL         | 0.770                             | 0.770        | 0.779        | 0.774        | 0.712        |

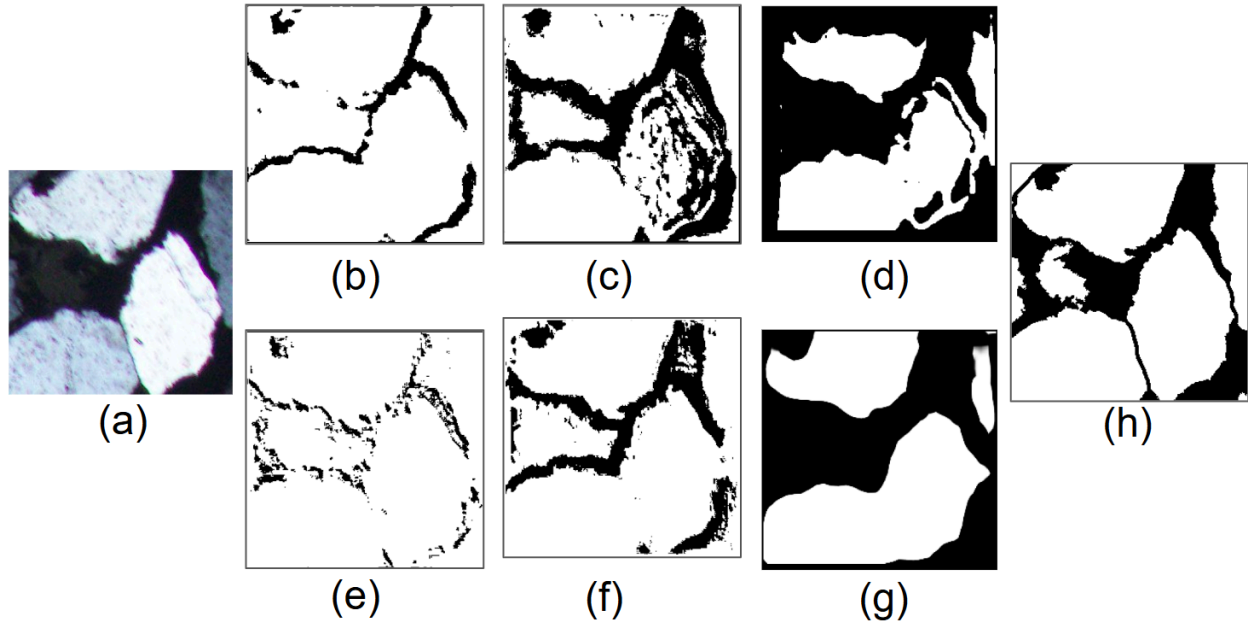
**Table 6.2:** Performance comparison among state of the art segmentation networks on Test-Set2. BCEL: binary cross entropy loss, DSL: Dice Score Loss, WBCEL: weighted binary cross entropy loss, and JI: Jaccard Index.

| Method  | Parameters |
|---------|------------|
| U-Net   | 30.00M     |
| U-Net++ | 9.04M      |
| FPNet   | 23.15M     |
| MANet   | 31.78M     |
| PSPNet  | 21.43M     |
| LinkNet | 11.50M     |

**Table 6.3:** Compares various architectures concerning parameters.

#### 6.4.4 Parameter Comparison

Table 6.3 highlights the comparison among semantic segmentation architectures with respect to their parameters. We observe from the table that MANet has the largest set of parameters (31.78M) among our experiments' considered networks. In contrast, U-Net++ has the smallest set of parameters (9.04M). LinkNet contains a moderate set of parameters which is considerably lesser than UNet and slightly larger than U-Net++. Training a network with a large number of parameters also requires a large number of training images. The same is not always available for real-world problems, e.g., the present case of grain segmentation. It is challenging and cost-intensive to create training images on grain segmentation. In the case of a lesser amount of training images, our method effectively solves the problem.



**Fig. 6.8:** Illustrates qualitative results obtained using various segmentation architectures. (a) Original XP image. (b) output of UNet. (c) output of UNet++. (d) output of FPNNet. (e) output of MANet. (f) output of LinkNet. (g) output of PSPNet. (h) Ground truth.

## 6.5 Conclusions

This chapter presents application of deep learning based methods in grain-matrix segmentation problem while making a two-fold contribution: first, instead of using only cross-polar images, as is usually done, we generate new images for training, by mixing cross-polar and plane-polar images through simple averaging, and second, we incorporate a new loss function, termed Weighted Binary Cross Entropy Loss, into the various deep learning models to assess their efficacy in automatically segmenting plane and cross-polarized microscopic sandstone images. We made six different training sets and two different test sets which include all the plane-polar, cross-polar and the newly generated images (combination of plane-polar and cross-polar). We compared our proposed loss function with two other well-known loss functions (Dice loss and Binary Cross Entropy Loss). Due to the unavailability of sufficient amounts of training images, we used various data augmentation schemes to circumvent the problem to some extent. The limited number of available images also motivates us to test another contemporary deep learning

method known as the self-supervised method discussed in Appendix C following Lin et al. (2023).

Our experimental results indicate that the majority of the cases from our SPGMS dataset yield superior outcomes when we implement our proposed loss function (WBCEL) into the deep learning based models. U-Net++ achieves the highest efficacy in all scenarios with the fewest parameters. The architecturally simpler LinkNet, which is also more lightweight than all other models in terms of number of parameters, except U-Net++, is interestingly found to be capable of generating results that are comparable to those of U-Net++ when equipped with WBCEL.

# Chapter 7

## Conclusion

### 7.1 Introduction

Automatic grain segmentation in clastic sedimentary rocks is a difficult and challenging job though it is highly desirable as it can lead to both time-efficient and labor-efficient real-time solution of this problem to the geologists. Though there have been some efforts, to date, for designing algorithms to this end, they actually fall short of expectations in performing the desired grain-matrix semantic segmentation. In the present scenario of research, not even a ready-made dataset exist that can be used for testing a new model or technique. To this end, in the present thesis, firstly a dataset consisting of sixteen different photomicrographs (both in plane polar and cross polar mode) has been built along with the manually developed ground truths for each of the cases, although the complete SPGMS dataset consists of 3 more images that have not been used for the work in the present thesis.

Secondly, we have shown by directly applying the state of the art segmentation algorithms that their performance is not up to the mark. On the other hand clustering based segmentation can significantly improve if preceded by a pre-processing step where we pass the respective images through suitable filters mimicking visual channels and mid-level vision in HVS. In the former (visual channel) case, the filter choice may be based on image patterns (e.g. grain size) while for the mid-level vision model, this choice is more automatic and adaptive. The psychophysics based methods used in both of these contexts are unsupervised in nature, requiring no training of the models by a-prior labeled images.

Thirdly, we have shown that semantic image segmentation can be considered as a machine learning task where the suitable features for segmentation can be hand-crafted i.e. trained from

proper datasets through supervised learning. As presently, the grains and matrix portions in the photomicrographs are very hard to discriminate due to highly overlapping, ambiguous features in them, we have investigated the applicability of some deep learning methods and applied them successfully with better results on our dataset utilizing the plane polar, the cross polar, and average images generated using both, dividing them into training set and testing set.

## **7.2 The Geological Significance of the present work**

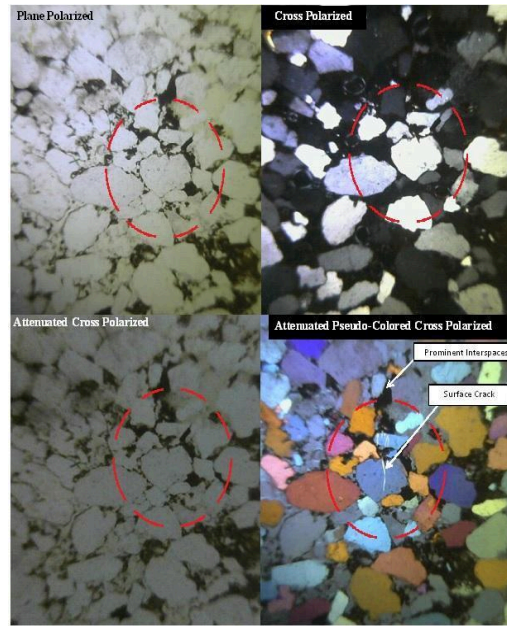
The framework grains in a sandstone sample, their sorting, roundness, distribution of the size population and the mineralogy are important for the interpretation of the sources, processes of deposition and their efficacy in hosting hydrocarbons or other economic mineral deposits [Friedman, 1961; Nag et al., 2016; Houben et al., 2020; Ji et al., 2022]. Traditionally these analyses are carried out by a geologist through the examination of thin sections of a sandstone under a petrographic microscope. The process is tedious, time consuming, often subjective. This thesis proceeds with the aim of automating the process of thin sections study by using a couple of thin section photomicrographs and subjecting it to various techniques of image analysis.

Most of the sandstone samples are characterized by widely varying sizes of the detrital framework grains, different mineralogy of these grains, varying nature of their mutual contact, and also varying proportions and types of matrix and cement that surround them. Thus, one of the major tasks for the automated granulometric analysis of the sandstone is to confidently distinguish between the framework grains and the matrix or cement surrounding them. In this work, a database, consisting of nineteen images was developed from the photomicrographs of a variety of sandstones so that different image processing techniques can be run on them.

## **7.3 Future Direction**

Successful grain-matrix segmentation using the presently developed database, as demonstrated in this thesis, indicates the potential of the vision inspired digital image processing and deep learning techniques in this sphere to further proceed to determine various parameters calculated by the geologist from the property of individual sand grains (like mean or modal size, sorting,

roundness or mineralogy of the sand grains). Such techniques can replace the traditional method of manual petrographic analysis of the sandstone with the automated techniques using images of the sandstone thin sections.



**Fig. 7.1** Thin rock section (50X magnification) under the lab-built polarizing microscope in all the modes, viz. normal PPL (top-left), normal XPL (top-right), attenuated (with white paper) XPL (left bottom), and attenuated (with transparent cellophane paper) XPL (right bottom). It can be clearly seen that only under the attenuated pseudo-color cross polarized mode using transparent cellophane sheets that a crackling has been identified unlike the other modes. The generated image in this mode is also a highly enhanced ideal version for image segmentation.

In Chapter 3, we mentioned some photomicrographs in the dataset that contain gray regions. These images have not been used in the present work. However, several real-life slides contain such ambiguous regions that are difficult to classify within the two classes mentioned in this thesis, even by the expert. Future work may be planned using Generative Adversarial Networks (GAN) to both generate and learn such realistic and more complex situations.

We have also developed a portable, low-cost, customized, and multi-modal digital polarizing microscope in our lab to enhance such complex images like these sandstone photomicrographs using transmission techniques as in the costly petrographic microscopes. The multiple modes of imaging of the lab-built microscope are – i) Plane polarized ii) Cross polarized iii) Attenuated cross polarized (White paper sheet) and iv) Attenuated pseudo-color cross polarized (Transparent Cellophane sheets) (Fig. 7.1). In case of the cross-polarized mode the edges of the grains are prominent and it appears that some of the grains are selectively more intensified in comparison to their surrounding ones. In the attenuated pseudo-colored mode each of the grains of the specimen shows multiple colors (which are a mixture of primary colors) and this facilitates the identification of cracks and different geometrical shapes present in the crystals of the specimen. In addition, the inter-spaces between the rocks are markedly prominent [Roy et al., 2022]. The formation of pseudo-coloration of the specimen structures in the presence of the transparent cellophane sheets is possibly due to the anisotropy and birefringence of the specimens, and the image thus generated in this mode is also a highly enhanced ideal version for image segmentation, comparable to similar images found in the recent work of Murtazin et al. (2023). More experimentation and automated segmentation using such images need to be planned at this stage to do justice to this developmental work. If it is found consistently successful, then such images can also be added to the dataset, in consultation with the geologists.

To conclude, in the future, newer algorithms and methodologies, both supervised and unsupervised, can be investigated for grain-matrix segmentation. Similarly, new or modified computational models mimicking lower and mid-level HVS can be tested on the dataset for better explainability of the petrographer's vision that was attempted here. These receptive fields may be combined with the deep neural nets for further improving the results in achieving important tasks like separation of the overlapped grains with higher accuracy and making a more realistic grain count, a highly demanding job to date, in general, for most cases of thin section samples.



# List of Publications

## Conference papers:

1. Das R., Shankar B.U., Chakraborty T., Ghosh K., (2020). Addressing Grain-Matrix Differentiation in Sedimentary Rock Photomicrographs in the Light of Brightness Perception Modeling. In: Mandal J., Mukhopadhyay S. (eds) Proceedings of the Global AI Congress 2019. Advances in Intelligent Systems and Computing, vol 1112. Springer, Singapore.
2. Das R., Ghosh K., (2022). Solving a difficult figure-ground segmentation problem in sedimentary rock photomicrographs using a fully automated algorithm combining Marr's Raw Primal Sketch and a Magno-Parvo Additive model. Proceedings of the 44th European Conference on Visual Perception-ECVP 2022, Nijmegen, Netherlands, August 28-September 1, 2022 (to be published in a special supplemental issue of the journal Perception).
3. Das, R., Chowdhury, A., Chakraborty, T., & Ghosh, K. (Accepted for SPIE Proceedings). *SPGMS: An Annotated Benchmark for Grain-Matrix Segmentation in Sandstone Photomicrographs. ICVIP 2024 (The 8th International Conference on Video and Image Processing) which will be held on December 13-15, 2024 in Kuala Lumpur, Malaysia.*

## Journal papers:

1. Das, R., Shankar, B. U., Chakraborty, T., & Ghosh, K. (2021). Automatic grain segmentation in cross-polarized photomicrographs of sedimentary rocks using psychophysics inspired models. Innovations in Systems and Software Engineering, 17(2), 167-183.
2. Das, R., Mondal, A., Chakraborty, T., & Ghosh, K. (2022). Deep neural networks for automatic grain-matrix segmentation in plane and cross-polarized sandstone photomicrographs. Applied Intelligence, 52(3), 2332-2345.

# Bibliography

- Abutaleb, A. S. (1989). Automatic Thresholding Of Gray-Level Pictures Using Two Dimensional Entropy. *Computer Vision,Graphics, And Image Processing*, 47, 22-32. 1989.
- Abutaleb, A., and A. Eloteifi (1988). "Automatic thresholding of gray-level pictures using 2-D entropy." *Applications of Digital Image Processing X*. Vol. 829. SPIE, 1988.
- Achanta, R., Shaji, A., Smith, K., Lucchi, A., Fua, P., & Süsstrunk, S. (2012). SLIC superpixels compared to state-of-the-art superpixel methods. *IEEE transactions on pattern analysis and machine intelligence*, 34(11), 2274-2282.
- Aljuaid, A., & Anwar, M. (2022). Survey of supervised learning for medical image processing. *SN Computer Science*, 3(4), 292.
- Altermann, W (2002). The evolution of life and its impact on sedimentation, In: *Precambrian sedimentary environments: a modern approach to ancient depositional systems*, 15–32.
- Arnab, A., Jayasumana, S., Zheng, S., & Torr, P. H. (2016, October). Higher order conditional random fields in deep neural networks. In the *European conference on computer vision* (pp. 524-540). Springer, Cham.
- Arthur D, Vassilvitskii S. (2006). k-means++: the advantages of careful seeding. Technical report, Stanford.
- Averill, Stuart A. (2001).The application of heavy indicator mineralogy in mineral exploration with emphasis on base metal indicators in glaciated metamorphic and plutonic terrains. Geological Society, London, Special Publications, 185, 69-81.
- Badrinarayanan, V., Kendall, A., & Cipolla, R. (2017). Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE transactions on pattern analysis and machine intelligence*, 39(12), 2481-2495.
- Barik, D. and Mondal, M.(2010), "Object identification for computer vision using image segmentation," in [2010 2nd International conference on education technology and computer], 2, V2–170, IEEE .

- Beyer, R. A., Stack, K. M., Griffes, J. L., Milliken, R. E., Herkenhoff, K. E., Byrne, S., Holt, J. W. & Grotzinger, J. P. (2012). An atlas of mars sedimentary rocks as seen by HiRISE, In: *Sedimentary Geology of Mars*, SEPM Special Publication, 102, 49-95.
- Blakeslee, B., & McCourt, M. E. (2004). A unified theory of brightness contrast and assimilation incorporating oriented multiscale spatial filtering and contrast normalization. *Vision research*, 44(21), 2483-2503.
- Blatt H. (1992). *Sedimentary Petrology*, xiv+ 514 pp. New York, Oxford: WH Freeman & Co. Ltd.
- Bowker D. O.1983. Suprathreshold spatiotemporal response characteristics of the human visual system. *J Opt Soc Am*, 73, 436–440.
- Boykov, Y., & Jolly, M. (2001). "Interactive graph cuts for optimal boundary & region segmentation of objects in N-D images." *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*.
- Budenny S, Pechezhertsev A, Bukharev A, Erofeev A, Mitrushing D, Belozerov B. (2017). *Image processing and machine learning of SPE Russian petroleum technology conference*, Moscow, Russia, ISBN: 978-1-61399-587-7.
- Bullier J. (2001). Integrated model of visual processing. *Brain Res Rev*, 36, 96-107.
- Canny, J. (1986) A computational approach to edge detection. *IEEE Trans Pattern Anal Mach Intell*, 8, 679–698.
- Caron, M., Bojanowski, P., Douze, M., Jégou, H., & Navarrete, J. (2018). "Deep clustering for unsupervised learning of visual features." *European Conference on Computer Vision (ECCV)*, 132-149.
- Chan K. (2018). Segmentation of moving objects in image sequence based on perceptual similarity of local texture and photometric features. *EURASIP Journal on Image and Video Processing* 2018 (1), 1-16.
- Chaurasia, A., & Culurciello, E. (2017, December). Linknet: Exploiting encoder representations for efficient semantic segmentation. In *2017 IEEE Visual Communications and Image Processing (VCIP)* (pp. 1-4). IEEE.
- Chayes F. (1949). A simple point counter for thin section analysis. *American Mineralogist*, 34, 1–11.

- Chen, L.C., Papandreou, G., Kokkinos, I., Murphy, K., Yuille, A.L. (2017). Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE transactions on pattern analysis and machine intelligence*, 40(4), 834-848.
- Chen, L. C., Yang, Y., Wang, J., Xu, W., & Yuille, A. L. (2016). Attention to scale: Scale-aware semantic image segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 3640-3649).
- Chen, X., & He, K. (2020). "A simple framework for contrastive learning of visual representations." *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 1597-1607.
- Chen, Y., Yi, Y., Dai, Y., and Shi, X., "A multiangle polarised imaging-based method for thin section segmentation," *Journal of Microscopy* (2024).
- Chow, C. K., & Kaneko, T. (1972). Automatic boundary detection of left ventricle from cineangiograms, *Comput. Biomed. Res.*, 5, 338-410.
- Chow, C. K., & Kaneko, T. (1972). Boundary detection of radiographic images by a threshold method. In *Frontiers of pattern recognition* (pp. 61-82). Academic Press.
- Comaniciu, D., & Meer, P. (2002). "Mean shift: A robust approach toward feature space analysis." *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(5), 603-619.
- Compton, R. R. (1962). Manual of field geology. *Soil Science*, 93(4), 295.
- Curtis, C. D. (1990). Aspects of climatic influence on the clay mineralogy and geochemistry of soils, palaeosols and clastic sedimentary rocks. *Journal of the Geological Society*, 147(2):351–357.
- Das, R., Shankar, B. U., Chakraborty, T., & Ghosh, K. (2020). Addressing grain-matrix differentiation in sedimentary rock photomicrographs in the light of brightness perception modelling. In *Proceedings of the Global AI Congress 2019* (pp. 223-235). Springer, Singapore.
- Das, R., Shankar, B. U., Chakraborty, T., & Ghosh, K. (2021). Automatic grain segmentation in cross-polarized photomicrographs of sedimentary rocks using psychophysics inspired models. *Innovations in Systems and Software Engineering*, 17(2), 167-183.

Das, R., Mondal, A., Chakraborty, T., & Ghosh, K. (2022). Deep neural networks for automatic grain-matrix segmentation in plane and cross-polarized sandstone photomicrographs. *Applied Intelligence*, 52(3), 2332-2345.

Das, R., Ghosh, K. 2022. Solving a difficult figure-ground segmentation problem in sedimentary rock photomicrographs using a fully automated algorithm combining Marr's Raw Primal Sketch and a Magno-Parvo Additive model. Proceedings of the 44th European Conference on Visual Perception-ECVP 2022, Nijmegen, Netherlands, August 28-September 1, 2022 (to be published in a special supplemental issue of the journal Perception).

*Das, R., Chowdhury, A., Chakraborty, T., & Ghosh, K. (Accepted for SPIE Proceedings). SPGMS: An Annotated Benchmark for Grain-Matrix Segmentation in Sandstone Photomicrographs. ICVIP 2024 (The 8th International Conference on Video and Image Processing) which will be held on December 13-15, 2024 in Kuala Lumpur, Malaysia.*

Delesse, M. A. (1848). Procédémécanique pour déterminer la compositiondesroches, Annales des Mines, 13, 379–388.

Deng, J., Dong, W., Socher, R., Li, L. J., Li, K., & Fei-Fei, L. (2009, June). Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition* (pp. 248-255). Ieee.

Desolneux, A., Moisan, L., & Morel, J. M. (2007). From gestalt theory to image analysis: a probabilistic approach (Vol. 34). Springer Science & Business Media.

De Valois RL, De Valois KK, 1988. Spatial Vision. New York: Oxford University Press.

De Weert CMM, Spillmann L. 1995. Assimilation: asymmetry between brightness and darkness. Vision Research 35:1413-1420.

Dias B, Ferdandez V, Victor L, Rodrigo S, Jose Y. (2020). Petrographic microscope digital image processing techniques for texture and microstructure interpretation of earth materials. Earth Sci Res 9(1):58.

Esmailzehi, Alireza, M. Omair Ahmad, and M. N. S. Swamy. "FPNet: A deep light-weight interpretable neural network using forward prediction filtering for efficient single image super resolution." IEEE Transactions on Circuits and Systems II: Express Briefs 69, no. 3 (2021): 1937-1941.

- Eser S.E.R.T., & Derya A.V.C.I. (2019). A new edge detection approach via neutrosophy based on maximum norm entropy. *Expert Systems with Applications*, 115, 499-511.
- Felzenszwalb, P.F., Huttenlocher, D.P. (2004). Efficient Graph-Based Image Segmentation. *International Journal of Computer Vision*, 59, 167–181.
- Fernando S.M.X. and Monro D.M. (1982). Variable thresholding applied to angiography, in *Proceedings, 6th International Conference on Pattern Recognition*.
- Folk, R. L. (1980). *Petrology of Sedimentary Rocks*. Austin: Hemphill Publishing.
- Fränti, P., & Sieranoja, S. (2019). How much can k-means be improved by using better initialization and repeats? *Pattern Recognition*, 93, 95-112.
- Friedman, G.M., and Sanders, J.E. (1978). *Principles of Sedimentology*. New York: Wiley.
- Garcia-Pedrero, Angel & García-Cervigón, Ana & Olano, José & García-Hidalgo, Miguel & Lillo, Mario & Gonzalo-Martin, Consuelo & Caetano-Sánchez, Cristina & Calderón Ramírez, Saúl. (2020). Convolutional neural networks for segmenting xylem vessels in stained cross-sectional images. *Neural Computing and Applications*. 32. 10.1007/s00521-019-04546-6.
- Garg, A. (2016). A Review on Image Segmentation Techniques. *International Journal of Recent Research Aspects (IJRRA)*, 53-55.
- Ghosh, K. (2012). A possible role and basis of visual pathway selection in brightness induction. *Seeing Perceiving*, 25(2), 179-212.
- Ghosh, K. (2016). A neural network based model of M and P LGN cells, 2016. In: *IEEE proceedings, bioinformatics and systems biology (BSB2016)*, International conference on, 1-5, Allahabad, India, 2016.
- Ghosh, K., & Chakraborty, T. (2022). Impact of human intervention structures on the rivers: An investigation of the spatiotemporal variation of grain size in the Tista River, eastern Himalayas. *Earth Surface Processes and Landforms*, 47(9), 2245-2265.
- Goodchild, J. S., & Fueten, F. (1998). Edge detection in petrographic images using the rotating polarizer stage. *Computers & Geosciences*, 24(8), 745-751.
- Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., ... & Bengio, Y. (2014). "Generative adversarial nets." *Advances in Neural Information Processing Systems (NeurIPS)*, 27.

- Han, K., Sheng, V. S., Song, Y., Liu, Y., Qiu, C., Ma, S., & Liu, Z. (2024). Deep semi-supervised learning for medical image segmentation: A review. *Expert Systems with Applications*, 123052.
- Haralick, R. M., Shanmugam, K., & Dinstein, I. (1973). "Textural features for image classification." *IEEE Transactions on Systems, Man, and Cybernetics*, 3(6), 610-621.
- Heilbronner, R. (2000). Automatic grain boundary detection and grain size analysis using polarization micrographs or orientation images. *J Struct Geol*, 22(7), 969-981.
- Helson, H. (1963) Studies of anomalous contrast and assimilation. *J Opt SocAm*, 53, 179–184.
- He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 770-778).
- He, S., Schomaker, L. (2019) Deepotsu: Document enhancement and binarization using iterative deep learning. *Pattern recognition*, 91, 379-390.
- He, P., Xu, Y., Jiang, F., Wang, G., Xiao, Z., and Zheng, C. (2024), "A rapid evaluation method of blasting effect based on optimized image segmentation algorithm and application in engineering," *Scientific Reports* 14(1), 4783 .
- Hochstein S, Ahissar M, (2002). View from the top: Hierarchies and reverse hierarchies in the visual system. *Neuron*, 36, 791-804.
- Hou, X., & Zhang, L. (2007, June). Saliency detection: A spectral residual approach. In *2007 IEEE Conference on computer vision and pattern recognition* (pp. 1-8). Ieee.
- Huang, L. K., & Wang, M. J. J. (1995). Image thresholding by minimizing the measures of fuzziness. *Pattern recognition*, 28(1), 41-51.
- Izadi H, Sadri J, Mehran N\_A. (2015). A new intelligent method for minerals segmentation in thin sections based on a novel incremental color clustering. *ComputGeosci*, 81, 38-52.
- Izadi H, Sadri J, Bayati M. (2017). An intelligent system for mineral in thin section based on a cascade approach. *ComputGeosci*, 99, 37-49.
- Izadi H, Sadri J, Hormozzade F, Fattahpour V. (2020). Altered mineral segmentation in thin sections using an incremental-dynamic clustering algorithm. *Eng Appl ArtifIntell* 90:103466.

- Jameson D, Hurvich LM, (1989). Essay concerning color constancy. *Ann Rev Psych*, 40, 1–22.
- Jawahar, C., Biswas, P.K., Ray, A. (1997). Investigations on fuzzy thresholding based on fuzzy clustering. *Pattern Recognition*, 30(10), 1605–1613.
- Jennings, B.R., and Parslow, K., (1988). Particle Size Measurement: The Equivalent Spherical Diameter. *Proceedings of the Royal Society of London Series A*, 419(1856): 137-149.
- Jiang, F., Gu, Q., Hao, H., Li, N., Wang, B., & Hu, X. (2018). A method for automatic grain segmentation of multi-angle cross-polarized microscopic images of sandstone. *Computers & Geosciences*, 115, 143-153.
- J.S. Weszka. (1978). A Survey of Threshold Selection Techniques, *Computer Graphics and Image Processing*, 7, 259-265.
- Jungmann M, Pape H, Wibkirchen P, Clauser C, Berlage T. (2014). Segmentation of thin section images from grain size analysis using region competition and edge weighted region merging. *ComputGeosci*, 72, pp 33- 48.
- J. Xu, & K. Lange. (2019). "Power k-means clustering," In *International Conference on Machine Learning*, pp. 6921- 6931.
- Kapur, J. N., Sahoo, P. K., & Wong, A. K. (1985). A new method for gray-level picture thresholding using the entropy of the histogram. *Computer vision, graphics, and image processing*, 29(3), 273-285.
- Kaur, D. and Kaur, Y (2014)., "Various image segmentation techniques: a review," *International Journal of Computer Science and Mobile Computing* 3(5), 809–814.
- Kaymak, Ç. and Uçar, A. (2019), "A brief survey and an application of semantic image segmentation for autonomous driving," *Handbook of Deep Learning Applications* , 161–200.
- Kass, M., Witkin, A., & Terzopoulos, D. (1988). Snakes: Active contour models. *International journal of computer vision*, 1(4), 321-331.
- Kingma, D. P., & Welling, M. (2013). "Auto-Encoding Variational Bayes." *International Conference on Learning Representations (ICLR)*.
- K. Ding, L. Xiao, & G. Weng, "Active contours driven by local pre-fitting energy for fast image segmentation," *Pattern Recognition Letters*, vol.104, pp.29-36, 2018.



- Khawairakpam, R. A. Hazarika, & D. Kandar. (2019). "Image segmentation by fuzzy Edge detection and region growing technique," In Proceedings of the Third International Conference on Microelectronics, Computing and Communication Systems, pp. 51-64.
- Kingdom, F. A. (2011). Lightness, brightness and transparency: A quarter century of new ideas, captivating demonstrations and unrelenting controversy. *Vision Research*, 51(7), 652-673.
- Kittler, J., and Illingworth, J. (1986). Minimum error thresholding. *Pattern recognition*, 19(1), 41-47.
- Kohonen, T. (1995). *Self-Organizing Maps* (2nd ed.). Springer-Verlag.
- Krumbein, W. C. (1934). "Size frequency distributions of sediments". *Journal of Sedimentary Petrology*.
- Krumbein, W.C., and Sloss, L.L. (1963). *Stratigraphy and Sedimentation*, 2nd edition. San Francisco: W. H. Freeman.
- Kumar, R. K. Agrawal, & P. Kumar. (2020). Bias-corrected intuitionistic fuzzy c-means with spatial neighbourhood information approach for human brain MRI image segmentation. *IEEE Transactions on Fuzzy Systems*.
- Lafferty, J., McCallum, A., Pereira, F.C. (2001). Conditional random fields: Probabilistic models for segmenting and labeling sequence data. In: ICML.
- Leeder, M. R. (2011). Tectonic sedimentology: sediment systems deciphering global to local tectonics. *Sedimentology*, 58(1), 2-56.
- Li, C.H., Lee, C. (1993). Minimum cross entropy thresholding. *Pattern recognition*, 26(4), 617–625.
- Li, Rui, Shunyi Zheng, Ce Zhang, Chenxi Duan, Jianlin Su, Libo Wang, and Peter M. Atkinson. "Multiattention network for semantic segmentation of fine-resolution remote sensing images." *IEEE Transactions on Geoscience and Remote Sensing* 60 (2021): 1-13.
- Lin, X., Li, C. T., Adams, S., Kouzani, A., Jiang, R., He, L., et al. (2023). Self-supervised leaf segmentation under complex lighting conditions. *Pattern recognition*. 135, 109021.
- Lloyd, S. P. (1982). "Least squares quantization in PCM." *IEEE Transactions on Information Theory*, 28(2), 129-137.
- Lopez, I. Gloria. (2017). Grain Size Analysis. *Encyclopedia of Earth Science Series Encyclopedia of Geoarchaeology* Allan S. Gilbert (Volume Ed.) Springer.

- Lu, B., Cui, M., Wang, Y. (2009). Automated grain boundary detection using the level set method. *Comput Geosci*, 35(2), 267-275.
- Lumbreras, F. and Serrat, J. (1996). Segmentation of petrographical images of marbles. *Computers & Geosciences*, 22(5), 547–558.
- Maitre, J., Kevin, Bouchard. L., Bedard P. (2019). Mineral grains recognition using computer Vision and machine learning. *Comput Geosci*, 130, 84-93.
- Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information*. W.H. Freeman and Company, New York.
- Marr, D. & Hildreth, E (1980). A theory of edge detection. *Proceedings of the Royal Society of London. Series B. Biological Sciences*, 207(1167): 187-217.
- Mazumdar, D., Mitra, S., Ghosh, K., and Bhaumik, K. (2021). Analyzing the patterns of spatial contrast discontinuities in natural images for robust edge detection. *Pattern Analysis and Applications* (24). Springer, DOI:10.1007/s10044-021- 00976-y.
- Maunsell, J. H., Nealey, T. A., & DePriest, D. D. (1990). Magnocellular and parvocellular contributions to responses in the middle temporal visual area (MT) of the macaque monkey. *Journal of Neuroscience*, 10(10), 3323-3334.
- MacQueen, J., (1967). Some methods for classification and analysis of multivariate observations. *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, 1(14), 281-297.
- McManus, J. (1988). Grain size determination and interpretation. In Tucker, M. E. (ed.), *Techniques in Sedimentology*. Oxford: Blackwell Scientific, pp. 63–85.
- Merigan, W. H., Maunsell, J. R. H. (1993). How parallel are the primate visual pathways? *Ann Rev Neurosci.*, 16, 369-402.
- Meyer, F. (1992). "Topographic distance and watershed lines." *Signal Processing*, 38(1), 113-125.
- Minaee, S., Boykov, Y. Y., Porikli, F., Plaza, A. J., Kehtarnavaz, N., & Terzopoulos, D. (2022). Image segmentation using deep learning: A survey. *IEEE transactions on pattern analysis and machine intelligence*, 44(7), 3523-3542 .
- Mingireanov Filho, I., Spina, T. V., Falcão, A. X., & Vidal, A. C. (2013). Segmentation of sandstone thin section images with separation of touching grains using optimum path forest operators. *Computers & geosciences*, 57, 146-157.

- Mokhles M., Fatai A., Mohammed M. (2019). Advances in rock petrography: image processing techniques for automated textural thin section analysis. In: Proceedings of SPE middle east oil and gas show and conference, Manama, Baharin.
- Murtazin, T., Kayumov, Z., Morozov, V., Akhmetov, R., Kolchugin, A., Tumakov, D., ... & Sudakov, V. (2023). High-precision algorithm for grain segmentation of thin sections by multi-angle optical-microscopic images. *Journal of Sedimentary Research*, 93(12), 932-944.
- Najman, Y. (2006). The detrital record of orogenesis: A review of approaches and techniques used in the Himalayan sedimentary basins. *Earth-Science Reviews*, 74(1-2), 1-72.
- Nesse, W. (2000), "Introduction to mineralogy: Oxford university press," New York , 346–349.
- Nichols (2009), G. Sedimentology and stratigraphy. John Wiley & Sons.
- Nirgude, R. & Jain, S. ( 2014) Color image segmentation with k- means clustering and dynamic region merging. *Journal of Science, Engineering & Technology: A Peer Reviewed National Journal*, 1(5):1–10.
- Noh, H., Hong, S., & Han, B. (2015). Learning deconvolution network for semantic segmentation. In *Proceedings of the IEEE international conference on computer vision* (pp. 1520-1528).
- N. S. M. Raja, S. L. Fernandes, N. Dey, S. C. Satapathy, & V. Rajinikanth, "Contrast enhanced medical MRI evaluation using Tsallis entropy and region growing segmentation," *Journal of Ambient Intelligence and Humanized Computing*, 1-12.
- Ojala, T., Pietikäinen, M., & Harwood, D. (1996). "A comparative study of texture measures with classification based on featured distributions." *Pattern Recognition*, 29(1), 51-59.
- Otsu, N. (1979). A threshold selection method from gray-level histograms. *IEEE Transactions on Systems, Man, and Cybernetics* 9(1): 62-66.
- Pal, S. K, Ghosh, A. and Shankar, B. U. (2000). Segmentation of remotely sensed images with fuzzy thresholding, and quantitative evaluation. *International Journal of Remote Sensing*, 21(11):2269–2300.

- Patil, D. D. and Deore, S. G. (2013), "Medical image segmentation: a review," *International Journal of Computer Science and Mobile Computing* 2(1), 22–27.
- Pietikäinen, M. (2010). Local binary patterns. *Scholarpedia*, 5(3), 9775.
- Pouladzadeh, P., Shirmohammadi, S. & Arici, T (2013). Intelligent SVM based food intake measurement system, In: *Proceedings of CIVEMSA: The IEEE International Conference on Computational Intelligence and Virtual Environments for Measurement Systems and Applications*, pp. 87–92.
- Pont-Tuset, J., & Marques, F. (2015). Supervised evaluation of image segmentation and object proposal techniques. *IEEE transactions on pattern analysis and machine intelligence*, 38(7), 1465-1478.
- Rai, C. S. and Hanson, K. E. (1986). Shear-wave birefringence: A laboratory study. In: *SEG Technical Program Expanded Abstracts 1986*, Society of Exploration Geophysicists, pp. 471–473.
- Ray, A. K. (2008). *Fossils in earth sciences*. PHI Learning Pvt. Ltd.
- Reading, H. G. (2009). *Sedimentary environments: processes, facies and stratigraphy*. John Wiley & Sons.
- Romera, E., Alvarez, J. M., Bergasa, L. M., & Arroyo, R. (2017). Erfnet: Efficient residual factorized convnet for real-time semantic segmentation. *IEEE Transactions on Intelligent Transportation Systems*, 19(1), 263-272.
- Ronneberger, O., Fischer, P., Brox, T. (2015) U-net: Convolutional networks for biomedical image segmentation. In: *International Conference on Medical image computing and computer-assisted intervention*, pp. 234-241. Springer
- Rosenfeld, A., & De La Torre, P. (1983). Histogram concavity analysis as an aid in threshold selection. *IEEE Transactions on Systems, Man, and Cybernetics*, (2), 231-235.
- Roy, S., Bhattacharya, B., Das, R., Bal, B., Haldar, R., Ghosh, K. (2022) Development of a customized and multi-modal polarizing microscope for imaging of samples ranging from biological specimens to rock specimens, In: *Proceedings of XLV Symposium of the Optical Society of India- COPaQ 2022*, 171-173, <https://iitr.ac.in/copaq2022/img/COPaQ2022Proceeding.pdf> (awarded best poster)
- Ruddiman, W. R. (2008). *Earth's climate: past and future* 2nd WH Freeman and Company.

Sahoo, P., Wilkins, C., Yeager, J., 1997. Threshold selection using Renyi's entropy. *Pattern Recognition* 30(1): 71-84.

Sarkar, A., Mukherjee, A. D., Sharma, S., Sengupta, T., Ram, F., Bera, M. K., ... & Juyal, N. (2020). New evidence of early Iron Age to Medieval settlements from the southern fringe of Thar Desert (western Great Rann of Kachchh), India: Implications to climate-culture co-evolution. *Archaeological Research in Asia*, 21, 100163.

Saxena, N, Day-Stirrat, Ruarri J., Hows, Amie, Hofmann, Ronny, Application of deep learning for semantic segmentation of sandstone thin sections, *Computers & Geosciences* 152: 104778, 2021.

S. Fan, Y. Sun, & P. Shui, "Region-merging method with texture pattern attention for SAR image segmentation," *IEEE Geoscience and Remote Sensing Letters*, vol.18, no.1, pp.112-116. 2020.

Shah, B. K., Kedia, V., Raut, R., Ansari, S., & Shroff, A. (2020). Evaluation and Comparative Study of Edge Detection Techniques. *IOSR Journal of Computer Engineering*, 22(5), 6-15.

Shi, J. and Malik, J. (1997), "Normalized cuts and image segmentation", *Proceedings of IEEE computer society conference on computer vision and pattern recognition*, 731–737, IEEE, 1997.

Shi, J., & Malik, J. (2000). "Normalized cuts and image segmentation." *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(8), 888-905.

Shrivastava, K. Gupta, N. & Sharma, N. "Medical image segmentation using modified k means clustering," *International Journal of Computer Applications*, vol.103, no.16, pp.12–16, 2014.

Singh, A., Thomsen, K. J., Sinha, R., Buylaert, J. P., Carter, A., Mark, D. F., ... & Gupta, S. (2017). Counter-intuitive influence of Himalayan river morphodynamics on Indus Civilisation urban settlements. *Nature communications*, 8(1), 1-14.

Stanley-Wood, N., and Lines, R.W.(1992). *Particle Size Analysis*. Cambridge: Royal Society Of Chemistry. Royal Society of Chemistry, Special Publication 102.

Syvitski, J. P. M. (2007). *Principles, Methods and Application of Particle Size Analysis*. Cambridge: Cambridge University Press.

- Shand S. J. (1916). A recording micrometer for rock analysis, *Journal of Geology*, 24: 394–403.
- Tabor N. J. & Myers, T. S. ( 2015) Paleosols as indicators of paleoenvironment and paleoclimate. *Ann. Rev. Earth Planet Sci*, 43, 11.1-11.29
- Török, P, Higdon, P. D., Wilson, Ts (1998). On the general properties of polarised light conventional and confocal microscopes. *Optics communications*, 148(4-6):300–315.
- Tucker, M. (2001) *Sedimentary Petrology*. 3rd Edition, Blackwell Science, Oxford, 262
- Tucker M (1988) *Ed. Techniques in Sedimentology*, Blackwell Sc. Pub
- Tulasigeri, C. and Irulappan, M. "An advanced thresholding algorithm for diagnosis of glaucoma in fundus images," 2016 IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology (RTEICT), 2016, pp. 1676-1680, doi: 10.1109/RTEICT.2016.7808118.
- Van den Berg, E., Meesters, A., Kenter, J., and Schlager, W. (2002). Automated separation of touching grains in digital images of thin sections, *Computers & geosciences*, 28(2), 179–190.
- Van den Bergh, M., Boix, X., Roig, G., de Capitani, B., Van Gool, L. (2012). SEEDS: Superpixels Extracted via Energy-Driven Sampling. In: Fitzgibbon, A., Lazebnik, S., Perona, P., Sato, Y., Schmid, C. (eds) *Computer Vision – ECCV 2012*. ECCV 2012. Lecture Notes in Computer Science, vol 7578. Springer, Berlin, Heidelberg.
- Vincent, L., Soille, P.: Watersheds in digital spaces: an efficient algorithm based on immersion simulations. *IEEE Transactions on Pattern Analysis & Machine Intelligence* pp. 583-598 (1991) 42.
- Vo, Q.N., Kim, S.H., Yang, H.J., Lee, G.: Binarization of degraded document images based on hierarchical deep supervised network. *Pattern Recognition* 74, 568–586 (2018)
- Vo QN, Kim SH, Yang HJ, Lee G. (2018). Binarization of degraded document images based on hierarchical deep supervised network. *Pattern Recogn* 74: 568-586.
- Wolfe JM, Oliva A, Horowitz TS, Butcher SJ, Bompas A. (2002). Segmentation of objects from backgrounds in visual search tasks. *Vision Research* 42(28):2985-3004.
- Wang, C., Li, P., Long, Q., Chen, H., Wang, P., Meng, Z., Wang, X., and Zhou, Y. (2024), Deep learning for refined lithology identification of sandstone microscopic images, *Minerals* 14(3), 275.

- W. Wang, Z. Li, J. Yue, & D. Li, "Image segmentation incorporating double-mask via graph cuts," *Computers & Electrical Engineering*, vol. 54, pp. 246-254, 2016.
- Wang Q, Gao J, Yuan Y (2017) Embedding structured contour and location prior in siamesed fully convolutional networks for road detection. *IEEE Trans Intell Transp Syst* 19:230–241.
- Wei H, Wang L, Wang S, Jiang y, Li J.(2020). A signal-processing neural model based on biological retina. *Electronics* 9(1):pp.35-45.
- Wei H, 2016. A bio-inspired integration method for object semantic representation. *Journal of Artificial Intelligence and Soft Computing*, 6:137-154.
- Xie, S., Girshick, R., & Dollar, P. (2021). "Exploring simple siamese representation learning." *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 1738-1747.
- Yen, J.C., Chang, F.J., Chang, S.: A new criterion for automatic multilevel thresholding. *IEEE Transactions on Image Processing* 4(3), 370–378 (1995)
- Yesiloglu-Gultekin N, Keceli AS, Sezer EA, Can AB, Gokceoglu C, Bayhan H. (2012). A computer program (TSecSoft) to determine mineral percentages using photographs obtained from thin sections. *ComputGeosci*46: pp 310-316.
- Wolfe, J.M., Oliva, A., Horowitz, T.S., Butcher, S.J., Bompas, A.: Segmentation of objects from backgrounds in visual search tasks. *Vision research* 42(28), 2985–3004 (2002)
- Zhao, H., Shi, J., Qi, X., Wang, X., Jia, J.: Pyramid scene parsing network. In: *CVPR*, pp. 2881-2890. IEEE (2017)
- Zheng, S., Jayasumana, S., Romera-Paredes, B., Vineet, V., Su, Z., Du, D., Huang, C., Torr, P.H.: Conditional Random Fields as recurrent neural networks. In: *ICCV* (2015)
- Zhou, Y., Starkey, J., and Mansinha, L., "Segmentation of petrographic images by integrating edge detection and region growing," *Computers & Geosciences* 30(8), 817–831 (2004).
- Zhao, H., Shi, J., Qi, X., Wang, X., and Jia, J., "Pyramid scene parsing network," in [Proceedings of the IEEE conference on computer vision and pattern recognition], 2881–2890 (2017)

Zhou, Z., Siddiquee, M.M.R., Tajbakhsh, N., Liang, J.: Unet++: A nested U-Net architecture for medical image segmentation. In: Deep learning in medical image analysis and multimodal learning for clinical decision support, pp. 3-11. Springer (2018)



## Appendix A

**Detailed results of applying the psychophysics inspired neuro-visual model based algorithm for grain matrix segmentation as proposed in chapter 4**

| Input image | Method      | Quantitative Scores |               |               |               |               | Grain size |
|-------------|-------------|---------------------|---------------|---------------|---------------|---------------|------------|
|             |             | Dice score          | Accuracy      | Recall        | Precision     | Jaccard Index |            |
| Image1      | Original    | 0.6966              | 0.6984        | 0.5940        | 0.8420        | 0.5346        | Small      |
|             | M filtered  | 0.7444              | 0.7351        | 0.6231        | <b>0.9243</b> | 0.5930        |            |
|             | P filtered  | <b>0.7537</b>       | <b>0.7429</b> | <b>0.6428</b> | 0.9108        | <b>0.6048</b> |            |
|             | PM filtered | 0.7462              | 0.7361        | 0.6347        | 0.9052        | 0.5953        |            |
|             |             |                     |               |               |               |               |            |
| Image2      | Original    | 0.6761              | 0.6339        | 0.5213        | <b>0.9616</b> | 0.5107        | Small      |
|             | M filtered  | 0.6666              | 0.6250        | 0.5114        | 0.9569        | 0.4999        |            |
|             | P filtered  | <b>0.6898</b>       | <b>0.6443</b> | <b>0.5394</b> | 0.9563        | <b>0.5264</b> |            |
|             | PM filtered | 0.6722              | 0.6297        | 0.5178        | 0.9575        | 0.5062        |            |
|             |             |                     |               |               |               |               |            |
| Image3      | Original    | <b>0.6910</b>       | <b>0.6867</b> | <b>0.5414</b> | <b>0.9549</b> | <b>0.5279</b> | Big        |
|             | M filtered  | 0.6836              | 0.6800        | 0.5342        | 0.9491        | 0.5193        |            |
|             | P filtered  | 0.6890              | 0.6842        | 0.5406        | 0.9498        | 0.5256        |            |
|             | PM filtered | 0.6607              | 0.6631        | 0.5069        | 0.9483        | 0.4933        |            |
|             |             |                     |               |               |               |               |            |

|        |             |               |               |               |               |               |       |
|--------|-------------|---------------|---------------|---------------|---------------|---------------|-------|
| Image4 | Original    | 0.6903        | 0.6444        | 0.5368        | 0.9667        | 0.5271        | Big   |
|        | M filtered  | 0.6900        | 0.6428        | 0.5385        | 0.9603        | 0.5267        |       |
|        | P filtered  | 0.6322        | 0.5970        | 0.4691        | <b>0.9691</b> | 0.4622        |       |
|        | PM filtered | <b>0.6976</b> | <b>0.6504</b> | <b>0.5461</b> | 0.9653        | <b>0.5356</b> |       |
|        |             |               |               |               |               |               |       |
| Image5 | Original    | <b>0.6399</b> | <b>0.6171</b> | <b>0.4771</b> | 0.9713        | <b>0.4705</b> | Big   |
|        | M filtered  | 0.5968        | 0.5864        | 0.4292        | <b>0.9791</b> | 0.4253        |       |
|        | P filtered  | 0.6382        | 0.6168        | 0.4739        | 0.9768        | 0.4687        |       |
|        | PM filtered | 0.6203        | 0.6028        | 0.4550        | 0.9744        | 0.4496        |       |
|        |             |               |               |               |               |               |       |
| Image6 | Original    | <b>0.7346</b> | <b>0.7582</b> | <b>0.6276</b> | 0.8855        | <b>0.5805</b> | Small |
|        | M filtered  | 0.7042        | 0.7412        | 0.5780        | <b>0.9009</b> | 0.5434        |       |
|        | P filtered  | 0.7165        | 0.7479        | 0.5976        | 0.8943        | 0.5582        |       |
|        | PM filtered | 0.7211        | 0.7504        | 0.6054        | 0.8917        | 0.5639        |       |
|        |             |               |               |               |               |               |       |
| Image7 | Original    | 0.5814        | 0.6298        | 0.4692        | 0.7640        | 0.4098        | Small |
|        | M filtered  | <b>0.6139</b> | <b>0.6428</b> | <b>0.5183</b> | 0.7527        | <b>0.4429</b> |       |
|        | P filtered  | 0.5903        | 0.6326        | 0.4830        | 0.7588        | 0.4187        |       |
|        | PM filtered | 0.5819        | 0.6339        | 0.4651        | <b>0.7772</b> | 0.4104        |       |
|        |             |               |               |               |               |               |       |
| Image8 | Original    | 0.7298        | 0.7837        | <b>0.6508</b> | 0.8306        | 0.5745        | Big   |
|        | M filtered  | 0.7392        | 0.7931        | 0.6192        | 0.9169        | 0.5863        |       |
|        | P filtered  | 0.7332        | 0.7872        | 0.6228        | 0.8912        | 0.5788        |       |

|         |             |        |        |        |        |        |       |
|---------|-------------|--------|--------|--------|--------|--------|-------|
|         | PM filtered | 0.7398 | 0.7888 | 0.6156 | 0.9268 | 0.5871 |       |
|         |             |        |        |        |        |        |       |
| Image9  | Original    | 0.7901 | 0.7682 | 0.6634 | 0.9766 | 0.6530 | Small |
|         | M filtered  | 0.8019 | 0.7776 | 0.6844 | 0.9681 | 0.6693 |       |
|         | P filtered  | 0.8150 | 0.7910 | 0.7002 | 0.9749 | 0.6878 |       |
|         | PM filtered | 0.8111 | 0.7857 | 0.6996 | 0.9649 | 0.6823 |       |
|         |             |        |        |        |        |        |       |
| Image10 | Original    | 0.7858 | 0.7829 | 0.6779 | 0.9346 | 0.6472 | Small |
|         | M filtered  | 0.7796 | 0.7753 | 0.6764 | 0.9200 | 0.6388 |       |
|         | P filtered  | 0.7799 | 0.7774 | 0.6712 | 0.9306 | 0.6392 |       |
|         | PM filtered | 0.7513 | 0.7554 | 0.6290 | 0.9328 | 0.6017 |       |
|         |             |        |        |        |        |        |       |
| Image11 | Original    | 0.6726 | 0.5363 | 0.5123 | 0.9788 | 0.5067 | Big   |
|         | M filtered  | 0.6917 | 0.5525 | 0.5399 | 0.9622 | 0.5287 |       |
|         | P filtered  | 0.6733 | 0.5364 | 0.5138 | 0.9764 | 0.5075 |       |
|         | PM filtered | 0.6926 | 0.5533 | 0.5413 | 0.9614 | 0.5298 |       |
|         |             |        |        |        |        |        |       |
| Image12 | Original    | 0.7683 | 0.8017 | 0.6797 | 0.8836 | 0.6238 | Small |
|         | M filtered  | 0.7662 | 0.7900 | 0.7109 | 0.8308 | 0.6210 |       |
|         | P filtered  | 0.7767 | 0.8026 | 0.7096 | 0.8579 | 0.6349 |       |
|         | PM filtered | 0.7675 | 0.7954 | 0.6978 | 0.8528 | 0.6227 |       |
|         |             |        |        |        |        |        |       |
| Image13 | Original    | 0.4954 | 0.6127 | 0.3401 | 0.9118 | 0.3292 | Small |

|         |             |               |               |               |               |               |                                         |
|---------|-------------|---------------|---------------|---------------|---------------|---------------|-----------------------------------------|
|         | M filtered  | <b>0.5681</b> | <b>0.6487</b> | <b>0.4133</b> | 0.9084        | <b>0.3968</b> |                                         |
|         | P filtered  | 0.4822        | 0.6059        | 0.3283        | 0.9076        | 0.3177        |                                         |
|         | PM filtered | 0.5485        | 0.6396        | 0.3916        | <b>0.9152</b> | 0.3779        |                                         |
|         |             |               |               |               |               |               |                                         |
| Image14 | Original    | 0.5991        | 0.6418        | 0.4978        | 0.7521        | 0.4276        | Small                                   |
|         | M filtered  | 0.6076        | 0.6473        | 0.5080        | 0.7559        | 0.4364        |                                         |
|         | P filtered  | 0.6057        | 0.6432        | <b>0.5098</b> | 0.7460        | 0.4344        |                                         |
|         | PM filtered | <b>0.6085</b> | <b>0.6482</b> | 0.5087        | <b>0.7572</b> | <b>0.4373</b> |                                         |
|         |             |               |               |               |               |               |                                         |
| Image15 | Original    | 0.6028        | 0.6195        | 0.4580        | 0.8815        | 0.4314        | Moderate<br>sized grains                |
|         | M filtered  | 0.6028        | 0.6201        | 0.4572        | 0.8844        | 0.4314        |                                         |
|         | P filtered  | 0.5991        | 0.6177        | 0.4531        | 0.8840        | 0.4277        |                                         |
|         | PM filtered | <b>0.6042</b> | <b>0.6213</b> | <b>0.4585</b> | <b>0.8857</b> | <b>0.4329</b> |                                         |
|         |             |               |               |               |               |               |                                         |
| Image16 | Original    | <b>0.6440</b> | <b>0.5997</b> | <b>0.5048</b> | 0.8892        | <b>0.4749</b> | Small grains<br>except one<br>big grain |
|         | M filtered  | 0.6403        | 0.5965        | 0.5007        | 0.8878        | <b>0.4709</b> |                                         |
|         | P filtered  | 0.6399        | 0.5973        | 0.4988        | 0.8923        | 0.4705        |                                         |
|         | PM filtered | 0.6192        | 0.5824        | 0.4734        | <b>0.8950</b> | 0.4485        |                                         |

**Table A.1:** Detailed results of applying the psychophysics inspired neuro-visual model based algorithm for grain matrix segmentation as proposed in chapter 4, with highest values highlighted in boldface.

## Appendix B

**Confusion matrices for binary pixel classification for different unsupervised segmentation algorithms**

| n<br>=4700160   |          | Actual class   |                |
|-----------------|----------|----------------|----------------|
|                 |          | Positive       | Negative       |
| Predicted class | Positive | TP<br>=1858374 | FP<br>=84962   |
|                 | Negative | FN<br>=1586826 | TN<br>=1169998 |

| n<br>=4700160   |          | Actual class   |                |
|-----------------|----------|----------------|----------------|
|                 |          | Positive       | Negative       |
| Predicted class | Positive | TP<br>=1804087 | FP<br>=80425   |
|                 | Negative | FN<br>=1641113 | TN<br>=1174535 |

| n<br>=4700160   |          | Actual class   |                |
|-----------------|----------|----------------|----------------|
|                 |          | Positive       | Negative       |
| Predicted class | Positive | TP<br>=1285903 | FP<br>=54293   |
|                 | Negative | FN<br>=2159297 | TN<br>=1200667 |

(a)

(b)

(c)

| n<br>=4700160   |          | Actual class   |               |
|-----------------|----------|----------------|---------------|
|                 |          | Positive       | Negative      |
| Predicted class | Positive | TP<br>=1817112 | FP<br>=711312 |
|                 | Negative | FN<br>=1628088 | TN<br>=543648 |

| n<br>=4700160   |          | Actual class   |                |
|-----------------|----------|----------------|----------------|
|                 |          | Positive       | Negative       |
| Predicted class | Positive | TP<br>=1480967 | FP<br>=49152   |
|                 | Negative | FN<br>=1964233 | TN<br>=1205808 |

| n<br>=4700160   |          | Actual class   |                |
|-----------------|----------|----------------|----------------|
|                 |          | Positive       | Negative       |
| Predicted class | Positive | TP<br>=1437862 | FP<br>=48920   |
|                 | Negative | FN<br>=2007338 | TN<br>=1206040 |

|                 |          |                |                |                 |          |                |                |                 |          |                |                |
|-----------------|----------|----------------|----------------|-----------------|----------|----------------|----------------|-----------------|----------|----------------|----------------|
| (d)             |          |                |                | (e)             |          |                |                | (f)             |          |                |                |
| n<br>=4700160   |          | Actual class   |                | n<br>=4700160   |          | Actual class   |                | n<br>=4700160   |          | Actual class   |                |
|                 |          | Positive       | Negative       |                 |          | Positive       | Negative       |                 |          | Positive       | Negative       |
| Predicted class | Positive | TP<br>=1587623 | FP<br>=61405   | Predicted class | Positive | TP<br>=2151431 | FP<br>=134267  | Predicted class | Positive | TP<br>=1599174 | FP<br>=62510   |
|                 | Negative | FN<br>=1857577 | TN<br>=1193555 |                 | Negative | FN<br>=1293769 | TN<br>=1120693 |                 | Negative | FN<br>=1846026 | TN<br>=1192450 |
| (g)             |          |                |                | (h)             |          |                |                | (i)             |          |                |                |
| n<br>=4700160   |          | Actual class   |                | n<br>=4700160   |          | Actual class   |                | n<br>=4700160   |          | Actual class   |                |
|                 |          | Positive       | Negative       |                 |          | Positive       | Negative       |                 |          | Positive       | Negative       |
| Predicted class | Positive | TP<br>=1436584 | FP<br>=47039   | Predicted class | Positive | TP<br>=1436584 | FP<br>=47039   | Predicted class | Positive | TP<br>=1436584 | FP<br>=47039   |
|                 | Negative | FN<br>=2008616 | TN<br>=1207921 |                 | Negative | FN<br>=2008616 | TN<br>=1207921 |                 | Negative | FN<br>=2008616 | TN<br>=1207921 |
| (j)             |          |                |                | (j)             |          |                |                | (j)             |          |                |                |

**Fig. B.1:** Confusion matrices for the unsupervised algorithms' (stated in this thesis) results on one typical cross-polar image from our built dataset (here 'n' denotes the total number of pixels in the input image and TP, FP, FN, TN denote the number of true positive, false positive, false negative and true negative pixels respectively), for: (a) our proposed psychophysics inspired neuro-visual model in Chapter 4 [Das et al., 2021], (b) our proposed mid-level vision model of adaptive scrutiny, in Chapter 5 [Das and Ghosh, 2022], with used gaussian variance=4, (c) Normalized Cut (NCut) [Shi and Malik, 1997] method, (d) Local Binary pattern (LBP) [Pietikäinen, 2010] method, (e) Otsu [Otsu et al., 1975] thresholding method, (f) Kapur's entropy [Kapur et al., 1985] method of thresholding, (g) Yen's entropy [Yen et al., 1995] method of thresholding, (h) Rosenfeld [Rosenfeld and Torre, 1983] thresholding method, (i) Li's entropy [Li and Lee, 1993] method of thresholding, (j) Abutaleb [Abutaleb, 1989] thresholding method.

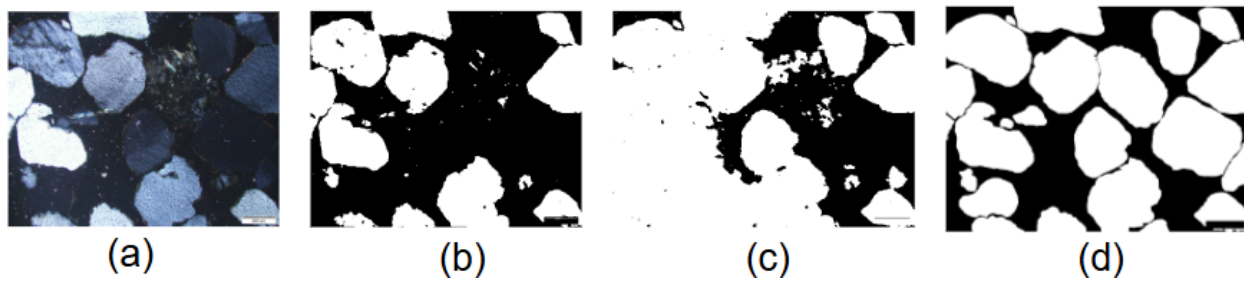
## Appendix C

### Results based on self-supervised segmentation

The self supervised segmentation method proposed in a very recent work in [Lin et al, 2023], used two of its important parameters viz. absolute threshold (at) and relative threshold (rt), for segmenting leaves under complex lighting conditions. Table C.1 shows the average segmentation performance for two variable sets of these two parameters' values. Here, it is worth mentioning that this method is stochastic in nature, thus may produce different output images for different trials of execution on the same input image, even for its fixed parameters' values. Figure C.1 shows output images of this self-supervised method for one input cross-polar image, using different sets of two of its most significant parameters' values.

| Two of its most significant<br>Parameter values | Image type | Average evaluation indices |        |           |       |       |
|-------------------------------------------------|------------|----------------------------|--------|-----------|-------|-------|
|                                                 |            | Accuracy                   | Recall | Precision | Dice  | Jl    |
| at=0.0095, rt=0.0                               | XP         | 0.551                      | 0.376  | 0.826     | 0.450 | 0.225 |
|                                                 | PP         | 0.478                      | 0.340  | 0.536     | 0.337 | 0.034 |
|                                                 | Combined   | 0.461                      | 0.230  | 0.733     | 0.267 | 0.116 |
| at=0.004, rt=0.0                                | XP         | 0.612                      | 0.621  | 0.790     | 0.635 | 0.238 |
|                                                 | PP         | 0.511                      | 0.468  | 0.587     | 0.457 | 0.037 |
|                                                 | Combined   | 0.523                      | 0.421  | 0.750     | 0.460 | 0.169 |

**Table C.1:** Average performance comparison of a self supervised method [Lin et al, 2022] recently used with two variable parameters set of values for xp, pp and combined corresponding to 16 triplets in our built dataset.



**Fig. C.1:** Output images under self-supervised method in [Lin et al, 2023] and ground truths for an input xp image. (a) Original xp mode input image, (b) Output image at  $at=0.0095$ ,  $rt=0.00$ , (c) Output image at  $at=0.004$ ,  $rt=0.00$  and (d) Ground truth image.