

Image Analysis and Machine Learning for Medical Images

by

Lina Liu

A thesis submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

in

Signal and Image Processing

Department of Electrical and Computer Engineering

University of Alberta

© Lina Liu, 2022

Abstract

Medical images play an essential role in detecting and diagnosing numerous diseases. With different medical imaging modalities, a large volume of images is generated every day across healthcare organizations worldwide, providing visualization of lesion appearance (e.g., dermoscopic images) anatomical information (e.g., MRI images), or cellular structures (e.g., laser scattering images). After obtaining the medical images, manual analysis and diagnosis is done by clinicians based on their prior knowledge and experiences. However, the diagnosis process can take a significant amount of time and the decision is subjective and biased toward different clinicians.

The main goal of this dissertation is to develop machine learning methods for automatic medical image analysis and diagnosis. In this dissertation, two types of medical images, including dermoscopic images and laser scattering images, are used. The dermoscopic images are used to detect the malignant lesions from the benign lesions. Experimental and simulated laser scattering images are used for label-free cell identification, and cell property characterization, respectively.

This thesis presents two methods based on machine learning for automatic skin lesion analysis. Skin lesion segmentation with auxiliary task is proposed for the accurate segmentation of the pigment regions, which does not require extra labeling information compared with the multi-task learning methods. An automatic skin lesion classification method based on mid-level feature learning is proposed for melanoma detection. State-of-the-art results have been

obtained and performances are discussed by extensive verification. This thesis further presents a machine learning technique for the analysis of laser scattering images. Typically, scattering patterns of the staurosporine-treated and non-treated SH-SY5Y neuroblastoma cells are obtained and classified, aiming at providing a better understanding of Parkinson's disease. In addition, multi-wavelength multi-direction laser scattering patterns of single cells have also been simulated to discuss the roles of two factors, cell surface roughness, and mitochondria number, in contributing to the scattering patterns. A systematic and thorough study has been done by extensive experiments. Theoretical analysis about the influence of the multi-wavelength multi-direction scattering patterns has been included. Satisfactory performance has been achieved for both the experimental and simulated data.

Preface

- Chapter 3 of this thesis has been published as Lina Liu, Lichao Mou, Xiao Xiang Zhu, and Mrinal Mandal. "Skin Lesion Segmentation based on improved U-net." In 2019 IEEE Canadian Conference of Electrical and Computer Engineering (CCECE), pp. 1-4, 2019. I was responsible for the proposed method, experiment design, code implementation, and manuscript composition. Lichao Mou and Prof. Xiao Xiang Zhu helped in the implementation process. Prof. Mrinal Mandal was the supervisory author and was involved with concept formation and manuscript composition. Meanwhile, another paper has been published as Lina Liu, Ying Yin Tsui, and Mrinal Mandal. "Skin lesion segmentation using deep learning with auxiliary task." Journal of Imaging, vol. 7, no. 4, pp. 67-82, 2021. I was responsible for the proposed method, experiment design, code implementation, and manuscript composition. Prof. Ying Yin Tsui helped in the manuscript composition. Prof. Mrinal Mandal was the supervisory author and was involved with concept formation and manuscript composition. This chapter proposes an automated technique to segment the pigment regions of dermoscopic images. Which is beneficial for the subsequent skin cancer (e.g., melanoma) detection.
- Chapter 4 of this thesis has been accepted in Computerized Medical Imaging and Graphics Journal as Lina Liu, Lichao Mou, Xiao Xiang Zhu, and Mrinal Mandal. "Automatic skin lesion classification based on mid-level feature learning." Computerized Medical Imaging and Graphics vol. 74, 2020. I was responsible for the proposed method and code implementation, signal processing, Prof. Lichao Mou and Prof. Xiao Xi-

ang Zhu helped in the manuscript composition process. Mrinal Mandal was the supervisory author and was involved with concept formation and manuscript composition. This chapter proposes an automated technique for skin lesion classification.

- Chapter 5 of this thesis has been published as Wendy Yu Wan, Lina Liu, Xiaoxuan Liu, Wei Wang, Md Zahurul Islam, Chunhua Dong, Craig R. Garen, Michael T. Woodside, Manisha Gupta, Mrinal Mandal, Wojciech Rozmus, and Ying Yin Tsui. "Integration of light scattering with machine learning for label free cell detection." *Biomedical Optics Express*, vol. 12, no. 6, pp. 3512-3529, 2021. I was the co-first author and was responsible for the machine learning technique, data analysis and code implementation. Wendy Yu Wan was responsible for the experiment setup and data collection. Wei Wang had contributed to the data simulation. Xiaoxuan Liu, Prof. Md Zahurul Islam, Prof. Wojciech Rozmus, and Prof. Mrinal Mandal helped in the manuscript composition. Chunhua Dong, Craig R. Garen, Michael T. Woodside helped in cell collection and processing. Prof. Ying Yin Tsui was the supervisory author and was involved with concept formation and manuscript composition. This chapter proposes a label-free technique to classify the non-treated and staurosporine-treated SH-SY5Y cells.
- Chapter 6 of this thesis will be submitted to a journal paper. I was responsible for the experiment design, code implementation, and manuscript composition. Prof. Md Zahurul Islam generated the simulated images. Xiaoxuan Liu helped in the data collection and manuscript composition process. Prof. Mandal helped in the method formulation and manuscript composition. Ying Yin Tsui was the supervisory author and was involved with the concept formation and manuscript composition. The technique uses machine learning method for the analysis of scattering patterns characterization regarding with the cell surface roughness and cell mitochondria number, which provides highlights for the future real experiment setup.

This dissertation is dedicated to my parents, who have been supporting and encouraging me over years.

Acknowledgements

Prof. Mrinal Mandal has been my supervisor for the last five years, and I would like to express my sincere gratitude towards him. Thanks for his patience, continuous support, and guidance during my graduate studies. I have benefited a lot from his knowledge and inspiration. Prof. Ying Yin Tsui has become my co-supervisor since the fourth year of the collaborative research with the label-free cytometry group. I learned and gained enormously from his experience and knowledge and would appreciate his support and valuable suggestions.

I would also like to thank the members of my defense committee, for taking their precious time to review my thesis and providing valuable suggestions. Much respect to the team members of the label-free cytometry, Wendy Yu Wan, Xiaoxuan Liu, Wei Wang, Prof. Md Zahurul Islam, Prof. Manisha Gupta, Prof. Wojciech Rozmus. Thanks for the meetings and all the conversations and discussions we have made. They have given me lots of inspiration and it is an enjoyable experience to work with them. Many thanks to my friends, who had make life at the University of Alberta colorfull. Thanks to the China Scholarship Council, who has awarded me the scholarship for studying at the University of Alberta.

Last but not least, thanks to my parents, my boyfriend Ruoheng Zhang, thanks for their continuous support, encouragement, careness, and love.

Contents

Abstract	ii
Preface	iv
Dedication	vi
Acknowledgements	vii
List of Tables	x
List of Figures	xii
Abbreviations	xviii
1 Introduction	1
1.1 Skin Cancer Detection and Dermoscopic Images	1
1.2 Label-free Cell Identification and Scattered Laser Light Patterns	5
1.3 Problem Statements and Motivations	8
1.4 Organization of the Thesis	10
2 Literature Review	12
2.1 Conventional Machine Learning and Deep Learning	12
2.2 Automatic Skin Lesion Segmentation	14
2.3 Automatic Skin Lesion Classification	17
2.4 Scattering Image Analysis Based on Machine Learning	20
2.5 Performance Evaluation Metrics	21
3 Skin Lesion Segmentation Using Deep Learning with Auxiliary Task	23
3.1 Proposed Method	25
3.1.1 Convolutional Neural Networks (CNN) Backbone	25
3.1.2 Cross-connection Layer (CCL)	27
3.1.3 Multi-scale Feature Aggregation (MSFA)	28
3.2 Dataset	31
3.3 Experimental Results	32
3.3.1 Implementation Details	32
3.3.2 Parameter Setting of the Loss Function	33
3.3.3 Ablation Study	34
3.3.4 Comparison with State-of-the-art Methods	35
3.4 Discussions	35
3.4.1 Quantitative Analysis of Test Results	35
3.4.2 Qualitative Analysis of Test Results	36
3.5 Conclusions	39
4 Automatic Skin Lesion Classification Based on Mid-level Feature Learning	40
4.1 Proposed Method	42
4.1.1 Skin Lesion Segmentation with Improved U-Net	43
4.1.2 Deep Feature Extraction	45

4.1.3	Mid-level Feature Learning	46
4.1.4	Classification Using SVM	50
4.2	Dataset	51
4.3	Experimental Results	51
4.3.1	Platform Information	51
4.3.2	Parameter Selection	51
4.3.3	Comparison with Features Extracted via Pretrained CNN	52
4.3.4	Comparison with Features Extracted via Finetuned CNN	54
4.3.5	Comparison with State-of-the-art Methods	56
4.4	Discussions	59
4.4.1	Segmentation Performance	59
4.4.2	Effect of Lesion Segmentation	60
4.4.3	Effect of Weighting Factor for Accelerated Proximal Gradient (APG) Algorithm	61
4.4.4	Advantage of Soft Discriminative Feature	62
4.4.5	Robustness of the Proposed Mid-level Features Against Parameters	63
4.4.6	Time Complexity	64
4.5	Conclusions	64
5	Integration of Light Scattering with Machine Learning for Label Free Cell Classification	65
5.1	Experiment Setup	67
5.2	Proposed Method	70
5.3	Dataset	71
5.4	Experimental Results	72
5.4.1	Platform Information	72
5.4.2	Comparison with Conventional Machine Learning Methods	72
5.4.3	Percentage Prediction in A Mixed Solution	74
5.5	Discussions	75
5.5.1	Comparison of the Proposed Technique Using Scattering Images From the Side Direction	75
5.5.2	Inference Time	76
5.6	Conclusions	76
6	Multi-wavelength Multi-direction Laser Light Scattering for Cell Characterization Using Machine Learning	77
6.1	Proposed Method	78
6.2	Data Generation	79
6.2.1	Numerical Simulation of Cell Surface Roughness	79
6.2.2	Numerical Simulation Models and the Angular Scattering Patterns	81
6.3	Dataset for Machine Learning	83
6.4	Experimental Results	85
6.4.1	Classification of Different Roughness Levels	85
6.4.2	Classification of Different Mitochondrial Numbers	88
6.5	Discussions	91
6.5.1	Quantitive Result Analysis	91
6.5.2	t-SNE Visualization	92
6.5.3	Insights Relevant to Previous Studies	94
6.6	Conclusions	95

7	Conclusions and Future Research Directions	97
7.1	Main Contributions	97
7.2	Future Research Directions	98
	References	100
	Appendix A Background Material	109
A.1	Cell Preparation in Chapter 5	109

List of Tables

3.1	Architecture of the proposed method. The input image size is 448×448	27
3.2	Experimental results with different values of α on the test set.	33
3.3	Ablation study of the proposed method	35
3.4	Experimental results compared with state-of-the-art methods on ISBI2017 test data.	35
4.1	Parameters used for performance evaluation.	52
4.2	Comparison of the proposed method with and without Mid-level Feature Learning (MFL) module.	52
4.3	Comparison of the proposed method with and without MFL module.	53
4.4	Comparison of the proposed method with the finetuned CNN and finetuned features using Area Under Curve (AUC) scores (in %).	55
4.5	Average AUC of the proposed method with different input sizes.	57
4.6	Ensemble performance with input of different scales on the validation set.	58
4.7	Performance comparison with state-of-the-art methods on ISIC 2017 dataset (AUC score).	58
4.8	Comparison of different methods on ISIC 2017 skin lesion segmentation dataset.	60
4.9	Average AUC of the proposed method with and without skin lesion segmentation.	60
4.10	Effect of weighting factor for APG algorithm.	61
4.11	Comparison of AUC scores of the proposed method using discriminative features and mid-level features.	62
5.1	Classification Performance of the proposed method.	74
5.2	Prediction of Percentage of the ST Cells (PTST).	75
5.3	Performance comparison of the proposed technique with scattering images from the side direction and forward direction.	76
6.1	Classification of scattering patterns with $r = 0.01$ and $r = 0.03$. The mitochondria number is 75. 5-fold cross validation is used, and the average performance is given.	86
6.2	Classification of scattering patterns with $r = 0.01$ and $r = 0.03$. The mitochondria number is 100. 5-fold cross validation is used, and the average performance is given.	87
6.3	Classification of scattering patterns with $r = 0.01$ and $r = 0.03$. For each group, an equal number of scattering patterns with 75 and 100 number of mitochondria are included. 5-fold cross validation is used, and the average performance is given.	87

6.4	Classification of scattering patterns with different numbers of mitochondria 75 and 100. The roughness value is $r = 0.01$. 5-fold cross validation is used, and the average performance is given.	89
6.5	Classification of scattering patterns with different numbers of mitochondria 75 and 100. The roughness value is $r = 0.03$. 5-fold cross validation is used, and the average performance is given.	90
6.6	Classification of scattering patterns with different numbers of mitochondria 75 and 100. For each group, an equal number of scattering patterns with $r = 0.01$ and $r = 0.03$ roughness values are included. 5-fold cross validation is used, and the average performance is given.	90

List of Figures

1.1	Layers of the skin and different types of skin cancers. Skin cancers are named for the type of cells where they start. Melanoma is the most aggressive skin cancer among different kinds of skin cancers (Image Credit: [15]).	2
1.2	Skin lesion examination using dermoscopy imaging. The dermoscopy images are of high-resolution, and are used as an essential adjunct to visual inspection of suspicious skin cancers (Image Credit: [58]).	3
1.3	Block diagram for automatic skin lesion analysis. The blue dashed box indicates the segmentation module, which may not exist in some methods. The yellow block represents a single CNN model, which can automatically learn the features and make predictions simultaneously.	3
1.4	Example images from ISIC 2017 dataset on skin lesion analysis towards melanoma detection. The images suffer from severe lighting condition variation, the existence of hairs, color marks, glue, and etc.	4
1.5	(a) (b) (c) are experimental and (g) (h) (i) simulated 2D laser scattered light patterns in the side direction for Platelet, Myeloid and CD 34+ Hematopoietic Stem Cell (HSC) respectively. (d) (e) (f) are their corresponding cell models used in the numerical simulation of 2D laser scattered light patterns of these cells. A red laser with a wavelength of 632.8 nm is used in the experiment.	6
2.1	Training of a model in machine learning.	12
2.2	Architecture of deep neural networks.	13
2.3	Typical architecture of CNN for a classification task. The current convolutional layers are obtained by convoluting previous layers with a set of 2-D convolution kernels. The pooling layers are obtained by subsampling the previous layers by a factor of 2.	15
2.4	Normal and abnormal skin lesions explained with the ABCDE rules. The abnormal skin lesion images are usually with asymmetric shape, irregular, jagged or blurred borders, multiple colors within the pigment regions, large diameter and evolving appearances over time.	18
3.1	Some skin lesion images from the ISBI2017 dataset for skin lesion segmentation. The ground truth masks are marked using the green contours. Fuzzy boundaries and noisy items such as hairs and color-marks are also witnessed.	25

3.2	Schematic diagram of the proposed method. Edge prediction is used as an auxiliary task to assist the segmentation task. Two parallel networks with Cross-connection Layer (CCL) module are implemented so that the learned two tasks can interact with each other during training and hence boost each other’s performance in turn. Multi-scale Feature Aggregation (MSFA) module is used to aggregate the multi-scale information from the intermediate feature maps of different scales.	26
3.3	Implementations of the CCL module. The inputs are feature maps F obtained from the backbone CNN. Fig. 3.3(a) shows the implementation details of a Residual block in Fig. 3.3(a), and blocks with the same color indicate the same operation. .	29
3.4	Implementations of the MSFA module, where the segmentation branch is used as an example. The inputs are feature maps of different scales from the segmentation branch, e.g., $S_{conv1}, S_{conv2}, S_{conv3}$. Four prediction masks are obtained after some convolution and pooling operation. The final output is weighted sum of these predictions which can be automatically learned by 1×1 convolution.	30
3.5	Experiment results with different values of α on the test set. .	34
3.6	Cumulative histograms of the Jaccard Index (JA) values on the test data.	36
3.7	Output visualization of the proposed method. (a) input test image; (b) the corresponding ground truth segmentation mask; (c) the output probability map of the segmentation prediction branch; (d) the output probability map of the edge prediction branch.	38
3.8	Some examples about the failed cases for the proposed method. (a) the input test images; (b) the corresponding ground truth segmentation masks; (c) the probability map of the proposed method.	39
4.1	Example images from ISIC 2017 dataset on skin lesion analysis towards melanoma detection. Each green box indicates a pair of inter-class samples (left: melanoma vs. right: benign), and each red box indicates a pair of intra-class samples (both images are melanoma). Strong inter-class visual similarity and intra-class variations are observed across different types of skin lesions. .	41
4.2	Block diagram of the proposed method.	42
4.3	Schematic of the proposed CNN based model for skin lesion segmentation. The proposed method contains an encoder network and decoder network, each network is consisted by a sequence of encoder blocks and decoder blocks.	44
4.4	The first row shows the original images, and the second row shows the images after pre-processing using Retinex method. The color distributions of different images are enhanced via the use of Retinex method. The resulted images are of similar lighting conditions.	45
4.5	Schematic plot of the proposed method.	48
4.6	t-distributed Stochastic Neighbor Embedding (t-SNE) visualization of the raw features extracted via pretrained ResNet and the learned mid-level features given raw features obtained from pretrained ResNet. The first and second row show scatter plots of the training data and testing data, respectively.	54

4.7	Examples of correctly classified images: left: melanoma; right: seborrheic keratosis.	59
4.8	Segmentation results of the proposed method. The red contours are the ground truths, and the blue contours are the segmentation results.	61
4.9	Performance of the proposed method (with and without the MFL module) on the parameter space. The learned mid-level features are more robust and discriminative compared with the original features.	63
5.1	Scanning Electron Microscopy (SEM) images of non-treated (left) and staurosporine-treated (right) SH-SY5Y cells. The two images correspond to cells of the similar size approximately 8 μm in diameter.	67
5.2	A schematic diagram of the experimental setup used to obtain two-dimensional light scattered patterns of cells in three directions.	68
5.3	Scattered light patterns from treated SH-SY5Y with red laser. The angular ranges corresponding to the three directions of measurements: 18°- 42° (forward), 79°-101° (side), and 141°-159° (backward).	69
5.4	Scattered light patterns from non-treated SH-SY5Y with red laser. The angular ranges corresponding to the three directions of measurements: 18°- 42° (forward), 79°-101° (side), and 141°-159° (backward).	69
5.5	Scattered light patterns from treated SH-SY5Y with green laser. The angular ranges corresponding to the three directions of measurements: 18°- 42° (forward), 79°-101° (side), and 141°-159° (backward).	70
5.6	Scattered light patterns from non-treated SH-SY5Y with green laser. The angular ranges corresponding to the three directions of measurements: 18°- 42° (forward), 79°-101° (side), and 141°-159° (backward).	70
5.7	Schematic of the machine learning module.	71
6.1	A cell cytoplasm model with the roughness model implemented. The following parameters are used: $\sigma = 0.9 * 520 \text{ nm}$, $\Lambda = \frac{2\pi r}{20}$ and $r = 5\text{nm}$, where r is the radius of the cell cytoplasm.	80
6.2	Example images of the cell models and their corresponding scattering patterns. The amplitudes of modulation for the cell surface roughness are 1%, 3%, 5% and 7% (sequentially from the left column to right column) of the the cell diameter (6.6 micron). Row 3-5 show the scattering patterns collected in the backward, side and forward direction. The wavelength is 432 nm.	82
6.3	Example of simulated scattering images collected in the forward direction.	84
6.4	Example of simulated scattering images collected in the side direction.	84
6.5	Example of simulated scattering images collected in the backward direction.	85
6.6	Scattering images with different roughness values r , which are collected in the backward direction. The mitochondria number is 100.	91

6.7	Example images with different mitochondria numbers N . The roughness value r is 0.01.	92
6.8	t-SNE visualization of light scattering patterns collected in the forward, side and backward directions with $\lambda = 432$ nm, 532 nm and 632 nm, and $N = 75$. The two groups of data with different roughness values r are denoted using different colours.	93
6.9	t-SNE visualization of scattering patterns collected in the forward, side and backward direction with wavelength $\lambda = 432$ nm, 532 nm and 632 nm. The roughness value is $r = 0.01$. The two groups of data with different mitochondrial number N are denoted using different colours.	94

Abbreviations

ABCDE Asymmetry, Border, Color, Diameter, Evolving.

ACC Accuracy.

APG Accelerated Proximal Gradient.

AUC Area Under Curve.

BCC Basal Cell Carcinoma.

BN Batch Normalization.

BP Bilinear Pooling.

CAD Computer-aided Diagnosis.

CAM Class Activation Mapping.

CCD Charged-coupled Device.

CCL Cross-connection Layer.

CNN Convolutional Neural Networks.

DC Dice Coefficient.

DL Deep Learning.

FACS Fluorescence-activated Cell Sorting.

FCM Fuzzy C-Means.

FCN Fully Convolutional Network.

FDTD Finite-Difference Time-Domain.

FPR False Positive Rate.

GAP Global Average Pooling.

GLCM Gray Level Co-occurrence Matrix.

GVF Gradient Vector Flow.

HOG Histogram of Oriented Gradient.

HSC Hematopoietic Stem Cell.

JA Jaccard Index.

LBP Local Binary Patterns.

MEL Melanoma.

MFL Mid-level Feature Learning.

MSFA Multi-scale Feature Aggregation.

NT None-treated.

PCA Principal Component Analysis.

PD Parkinson’s Disease.

PML Perfectly Matched Layer.

PPM Pyramid Pooling Module.

PTST Percentage of the ST Cells.

RBF Radial Basis Function.

RCNN Region Based Convolutional Neural Networks.

RELU Rectified Linear Unit.

ROI Regions of Interest.

SCC Squamous Cell Carcinoma.

SEM Scanning Electron Microscopy.

SEN Sensitivity.

SF Speckle Features.

SIFT Scale Invariant Feature Transformations.

SK Seborrheic Keratosis.

SPE Specificity.

ST Staurosporine-treated.

SVM Support Vector Machine.

TPR True Positive Rate.

t-SNE t-distributed Stochastic Neighbor Embedding.

VSEL Very Small Embryonic-like.

Chapter 1

Introduction

Computer-aided Diagnosis (CAD) systems are becoming increasingly important in diagnosing a wide range of diseases. The purpose of this thesis is to propose a methodology for CAD systems using medical images. Machine learning methods will be used for automatic analysis. Especially, two types of medical images: dermoscopic images and scattering images are considered in this thesis. The dermoscopic images are high-resolution RGB images that have been used as a primary step for skin cancer detection. On the other hand, laser scattering images are gaining popularity for cell identification in label-free cytometry application. In this chapter, a brief background of the dermoscopic images and scattering images will be introduced.

1.1 Skin Cancer Detection and Dermoscopic Images

The skin is the largest organ of human body, which is composed of three layers of tissue: the epidermis, dermis, and hypodermis. When the cells become disordered and grow out of control, they can develop into skin cancers and may spread to other body parts. Skin cancers are named based on the type of cells where they start. Melanocyte is a type of cell that is scattered in the area where the epidermis meets the dermis. It can produce melanin, the pigment that colors our skin. Skin cancer that starts in melanocytes is called melanoma, and those not are named as non-melanoma since they act very differently from melanomas. The two most common types of non-melanoma

are Basal Cell Carcinoma (BCC) and Squamous Cell Carcinoma (SCC). The schematic plot of the skin tissues and three types of skin cancers are shown in Fig. 1.1. There are sldo other types of unusual skin cancers, such as the Merkel cell tumor, actinic keratosis, and dermatofibrosarcoma protuberans.

Among all the different kinds of skin cancers, melanoma is the most aggressive one, whose incidence has increased rapidly over the past 30 years [72]. About 20% of Americans will get skin cancer during their lifetime [21]. Early detection is the best way to deal with melanoma since it is highly curable before spreading into other body parts. Specifically, the 5-year relative survival rate is 98% for the localized stage. To detect melanoma or suspected skin lesions, dermoscopy imaging is used to detect the pigmented skin lesions. It is a non-invasive technique and is used worldwide as a primary step. The dermoscopic images are with high resolution and enhanced visualization ability, which allow dermatologists to examine the skin lesions with naked eyes. Fig. 1.1 shows the process of skin lesion examination. A patient would be asked for further examination using the histopathological imaging if he or she has a high risk of skin cancer after analyzing his or her dermoscopic images.

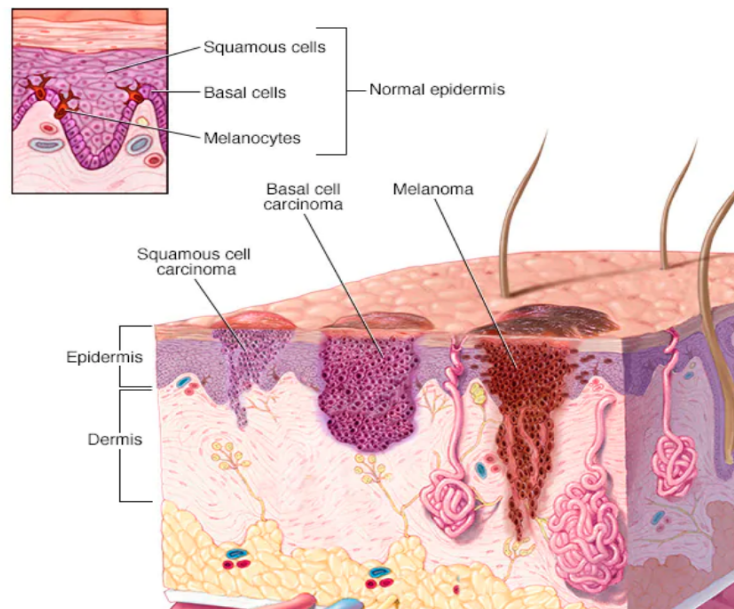


Figure 1.1: Layers of the skin and different types of skin cancers. Skin cancers are named for the type of cells where they start. Melanoma is the most aggressive skin cancer among different kinds of skin cancers (Image Credit: [15]).

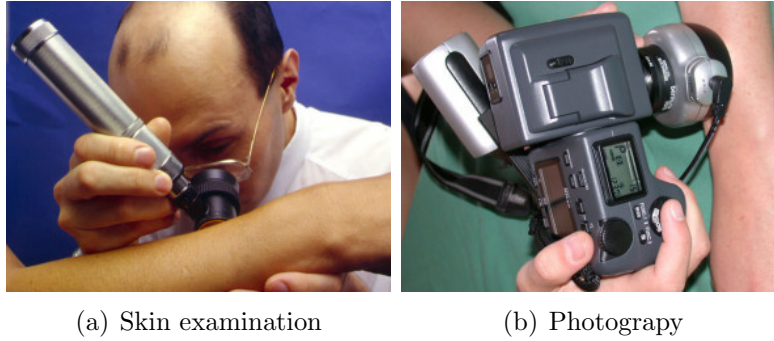


Figure 1.2: Skin lesion examination using dermoscopy imaging. The dermoscopy images are of high-resolution, and are used as an essential adjunct to visual inspection of suspicious skin cancers (Image Credit: [58]).

The dermoscopic images are of high resolution. Therefore, subpatterns and tiny structures can be visualized. The general pipeline for existing automated methods follows three steps: preprocessing, feature extraction, and classification. Skin lesion segmentation may also be performed before the feature extraction step to provide the boundary information or the Regions of Interest (ROI) to assist the subsequent classification task. Fig. 1.3 shows the block diagram for automatic skin lesion analysis. Notice that the yellow block, which includes three modules (segmentation, feature extraction, and classification) can be implemented using a convolutional neural network (CNN) model.

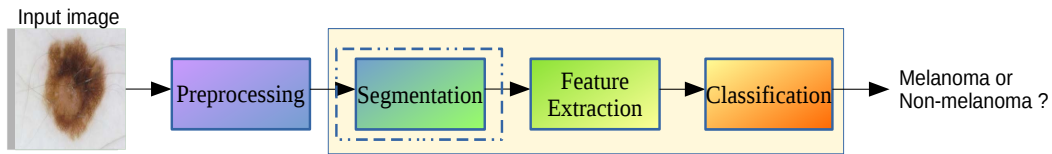


Figure 1.3: Block diagram for automatic skin lesion analysis. The blue dashed box indicates the segmentation module, which may not exist in some methods. The yellow block represents a single CNN model, which can automatically learn the features and make predictions simultaneously.

Fig. 1.4 shows some image samples from the ISIC 2017 dataset for skin lesion analysis, which is the most difficult dataset for skin lesion image analysis. The ISIC 2017 skin lesion analysis dataset includes three types of skin lesion images: the nevus and seborrheic keratosis, which are benign skin lesions; and melanoma, which is the most aggressive skin cancer. The images

present different lighting conditions and characteristics since they are collected across multiple institutes and hospitals using different photography instruments, which can reflect real cases mostly. As shown in Fig. 1.4, regardless of the strong lighting condition change, disturbing items such as hairs, color marks, rulers, and glue are also observed in the images. Therefore, the skin conditions are very complex, which further adds to the difficulty for automatic melanoma detection techniques. To address this problem, preprocessing techniques, such as color-enhancing [19], [20], hair removal [13], [37], [83] have been proposed in order to get clean images for the machine learning methods. However, such preprocessing methods rely heavily on prior knowledge, and may result in degenerated images with information loss, which can harm classification performance. Hence, the majority of machine learning methods focus on proposing more advanced models to deal with complex data. What is more, strong intra-class differences and inter-class similarities are also observed among different types of skin cancers.



Figure 1.4: Example images from ISIC 2017 dataset on skin lesion analysis towards melanoma detection. The images suffer from severe lighting condition variation, the existence of hairs, color marks, glue, and etc.

In summary, skin lesion analysis is a very difficult task, and the main challenges for this can be summarized as follows:

- Skin lesion images corresponding to the same type of cancer can present different visual patterns with dramatic appearance changes, while skin lesion images of different types of skin cancers can share strong visual

similarities. Therefore, large intra-class differences and inter-class similarities are observed for skin lesion images. For some difficult cases, it is even difficult for dermatologists to make decisions.

- For some skin lesions with light coloring, visual patterns and color of the pigment and skin regions are similar, with fuzzy and unclear boundaries, making it extremely difficult for the skin lesion segmentation task.
- The original dermoscopic images are usually of high resolution, which allows the observation of fine-grained structures and subpatterns. However, it is intractable for the computer to process the original image due to the large computation required for the high-resolution images.
- The original dermoscopic images suffer from severe lighting condition changes. In addition, the existence of disturbing items, such as hairs, color marks, and glues, may also deteriorate the learning performance.

1.2 Label-free Cell Identification and Scattered Laser Light Patterns

Another type of medical image, scattering patterns, is also considered in this PhD research. Scattering patterns can be used in label-free flow cytometry application. In this section, we will give a brief introduction to the label-free cell identification technique and the scattering patterns.

Cell identification plays a crucial role in many biology- and health-related applications. Conventional Fluorescence-activated Cell Sorting (FACS) flow cytometry is the current tool for cell identification. It is based on the measurement of fluorescence emissions of biomarkers that attach to a target cell when it is illuminated [62]. However, these biomarkers can interfere with the function of the cells they bind to, hindering further analysis of these cells [5]. Additionally, adequate biomarkers are not available for all cell types, and can be difficult and expensive to develop for new cell types. For example, the extraction of the newly discovered rare very Very Small Embryonic-like (VSEL) stem cells requires multiple steps and more than half a dozen biomarkers [63],

[98] led to significant loss and low purity. In addition, labeling requires considerable technical skills and often leads to inconsistent results.

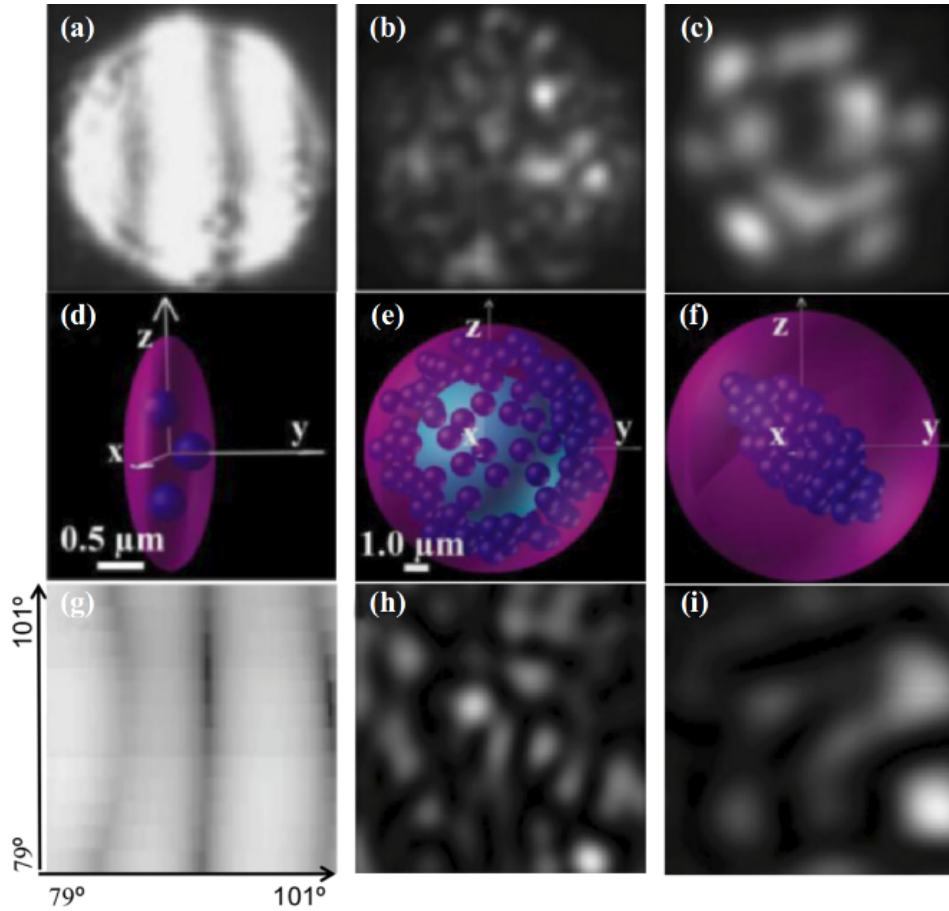


Figure 1.5: (a) (b) (c) are experimental and (g) (h) (i) simulated 2D laser scattered light patterns in the side direction for Platelet, Myeloid and CD 34+ HSC respectively. (d) (e) (f) are their corresponding cell models used in the numerical simulation of 2D laser scattered light patterns of these cells. A red laser with a wavelength of 632.8 nm is used in the experiment.

Cells have different properties in their shapes, sizes, number and distribution of the internal structures, etc. Variations of these properties will result in variations of the amplitude of scattered light in different directions. In other words, different scattering patterns can be obtained due to the variation of the cell properties. Therefore, the scattering patterns can characterize inner cellular structures and distributions within cells and have been studied as a label-free technique for single-cell analysis. The 2D laser light scattering patterns of single cells including a platelet cell, a myeloid precursor cell, a CD34+

HSC, in the side direction (polar angle of $70^\circ - 101^\circ$ respective to the laser axis) are shown in Fig. 1.5 [79]. The top and bottom row of images show the experimentally measured and computer-simulated scattering patterns, respectively. The second row of images are the cell models for the numerical simulation of 2D laser scattered light patterns. The 2D laser light scattered patterns can be simulated by solving the Maxwell Equation with relatively simple numerical models of cells that contain the main scattering components, such as the nucleus and mitochondria, and by describing non-spherical shape and surface irregularities. The platelet cell is assumed to be ellipsoidal with a minor axis of $1 \mu\text{m}$ and a major axis of $3 \mu\text{m}$ and contains only a small number of mitochondria as shown in Fig. 1.5 (a). The refractive index of the mitochondrion is assumed to be 1.42 inside a cell with a refractive index of 1.38. The myeloid precursor cell is assumed to be a $10 \mu\text{m}$ diameter sphere with a $6 \mu\text{m}$ diameter nucleus (in cyan) located at the center of the cell, as shown in Fig 1.5 (b). There are also 120 spherical mitochondria arbitrarily distributed inside the cell, each with a diameter of $1 \mu\text{m}$. The refractive indices for the myeloid precursor cell cytoplasm, nucleus, and mitochondria are set as 1.35, 1.39, and 1.42, respectively. The CD34+ HSC is also assumed to be a $10 \mu\text{m}$ diameter sphere containing mitochondria with aggregated distribution as shown in Fig. 1.5 (c) [79]. There are 70 mitochondria (diameter is set as $1 \mu\text{m}$), arbitrarily distributed in an ellipsoid centered at the origin with a major axis of $8 \mu\text{m}$ and a minor axis of $4 \mu\text{m}$. The refractive indices for the CD34+ HSC cytoplasm, mitochondria, and surrounding medium are assumed to be 1.35, 1.42, and 1.334, respectively. The laser beam is assumed to be a plane wave with an appropriate frequency. Both the experimental and simulated patterns of platelet cell show fringe structure as would be expected from a dielectric object lack of internal structures [79]. Both the experimental and simulated patterns of myeloid precursor cell and CD34+ HSC are dominated by the small-scale 2D speckle patterns originating from light scattering caused by mitochondria in the cells [78], [79]. The difference in these 2D laser light scattered patterns can be considered a signature for use in cell identification. For example, a relatively simple analysis using two observables in the 2D scat-

tered laser light patterns, the number of speckles and the average area of their cross-sections, are able to distinguish several blood cell types successfully [78], [79].

We have also studied two groups of SH-SY5Y neuroblastoma cells, including the Staurosporine-treated (ST) and None-treated (NT) neuroblastoma cells, using our laser light scattering technique to characterize the change of state of the cells when they are interacting with staurosporine. Death of dopamine-related neurons is one of several pathological indicators of Parkinson's Disease (PD); therefore, SH-SY5Y neuroblastoma cells have been widely used in PD research. Staurosporine is a natural substance that has been reported to induce SH-SY5Y cell death [57], while others have reported using staurosporine for SH-SY5Y differentiation into mature, neuron-like phenotypes [8]. The interaction of staurosporine with SH-SY5Y cells is therefore of interest to investigate whether a particular treatment dose will result in cell death or differentiation, which may benefit PD research. The scattered laser light patterns of the two groups of SH-SY5Y neuroblastoma cells look very similar. The simple two parameters method [78] used successfully for several blood cells identification is not sufficient to distinguish these two groups of cells from each other and hence there is a need for the development of machine learning analysis techniques.

1.3 Problem Statements and Motivations

We grade our PhD research into three applications, including skin lesion segmentation, skin lesion classification, cell classification and characterization. Prevalent machine learning methods for computer vision can be roughly grouped into the shallow learning method (conventional method) and the Deep Learning (DL) method. The DL method has become very popular since it can automatically learn the task-oriented optimal features and obtain state-of-the-art performance from the raw input data with little or no preprocessing, and has gained great interest in various research fields. With the encouraging results, deep neural networks have been regarded as the key method for future

applications in the industry domain. Among the neural networks of different architectures, the CNN has been widely used in computer vision tasks for image and video analysis. In this PhD thesis, I will focus on medical image analysis based on CNN.

For skin lesion segmentation task, despite the current research achievement, it is still a challenging task due to the following reasons: 1) For some skin lesions with light pigment, the color and visual patterns of the pigment regions and the skin regions are similar, with fuzzy and unclear boundaries, make it extremely difficult for the skin lesion segmentation task. 2) Complex skin conditions, including color inconsistency and disturbing items, such as hairs, veins, color marks, and other artifacts are also observed in the skin lesion images. Prevalent machine learning methods ensemble different CNN backbones to get the final performance, ignoring the intrinsic characteristics and auxiliary information of skin lesion images, which may benefit segmentation. Based on the aforementioned observation, a novel method based on the auxiliary task will be explored, which can segment the ROI accurately.

For the skin lesion classification task, strong intra-class difference and inter-class similarity is observed across different types of skin lesion. Dermatologists have to focus on the subtlety of details in order to distinguish malignant cases from benign ones, yet the large inter-class similarity and intra-class variations make it more formidable. In addition, the existence of complex skin conditions may introduce noisy items which can affect the color and texture description of a given image and deteriorate classification performance. Most CNN based methods for skin lesion analysis use different backbone structures, ignoring the discriminative power of the proposed model, which may lead to poor performance. They are also suffering from limitations such as lack of use of side information, no interpretability, etc. To address the aforementioned problems, the proposed methods should be discriminative, representative, and robust for various types of skin lesion images. The key idea is to learn an optimal feature representation, which satisfies the aforementioned characteristics. Novel methods regarding the aforementioned issues will be proposed.

For cell classification task, previous label-free cytometry technique col-

lected scattering patterns from one direction using one wavelength laser, and few machine learning methods have been proposed for scattering image analysis. In our work, scattering patterns are collected in three directions simultaneously, with red and green laser. Different from the convention RGB images from the digital camera, the laser scattering images are grayscale images, which share similar appearances and are difficult to discriminate with human eyes. A previous study has shown that using statistical analysis of the spatial distribution of the speckles in the scattered light patterns allows us to effectively distinguish one type of cell from the other. However, it has been found insufficient to use statistical features to discriminate between the non-treated and treated cells. Therefore, it is interesting to explore machine learning methods so as to classify these two types of cells. Conventional machine learning methods with hand-crafted features as well as CNN methods will be explored during the PhD period in order to make a thorough analysis. Since the data is small in size, transfer learning based on CNN will be used.

Based on the previous research on cell classification, the proposed machine learning technique will also be used for cell property characterization. Our motivation is that the internal and surface structures of a cell are complex and many features are sub-micron in scale, which cannot be obtained using optical microscopy. However, scattering patterns are affected by the cell properties and can be used as a potential way for cell characterization. Scattering patterns can also be affected by the laser wavelength and light collection direction. A systematic study about two cell properties (cell surface roughness and the number of mitochondria), probing wavelength and observation direction is conducted. The optimum wavelength and observation direction for each cell property are determined using extensive experiments.

1.4 Organization of the Thesis

The rest of the thesis is organized as follows: Chapter 2 presents a broad literature review of the prevalent machine learning methods for the dermoscopic image and scattering image analysis. In chapter 3, we first introduce

the benchmark CNN based method for skin lesion segmentation. An improved CNN model with auxiliary task learning is then proposed for accurate skin lesion segmentation. Chapter 4 introduces the skin lesion classification method based on mid-level feature learning. In Chapter 5, transfer learning is used for extracting features from real scattering images. A classifier is trained for cell classification. In Chapter 6, synthetic scattering images are generated with different cell properties (cell surface roughness, and mitochondria number) using multi-color multi-direction lasers. Machine learning methods are then used to analyze the role of these two factors in contributing to different scattering patterns. Finally, summaries and future work are introduced in chapter 7.

Chapter 2

Literature Review

In this section, we present a brief overview of the machine learning techniques for dermoscopic images and laser scattering images. The literature review includes a range of well-used image processing algorithms for ROI segmentation (skin lesion segmentation), skin cancer detection (skin lesion classification) and label-free cell classification. To better understand the machine learning methods, we first introduce two groups of methods of machine learning, the conventional machine learning methods and the DL methods.

2.1 Conventional Machine Learning and Deep Learning

Prevalent machine learning methods can be roughly grouped into the conventional machine learning methods (shallow learning method) and the DL methods.

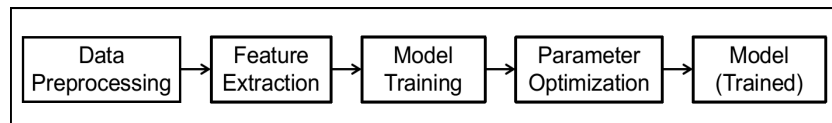


Figure 2.1: Training of a model in machine learning.

The conventional machine learning methods learn a trained model by using the training set as shown in Fig. 2.1. The learned model is typical a shallow model compared with the DL methods, where multiple hidden layer architectures are used to model the complex relationships between the input

and output. Fig. 2.2 shows a simple diagram of conventional DL methods, where the previous layers, after some calculation or operation, are fed into the later layers. The hidden layers can include conventional convolution layers, activation layers, pooling layers, etc. The networks can be very deep and up to several thousands hidden layers.

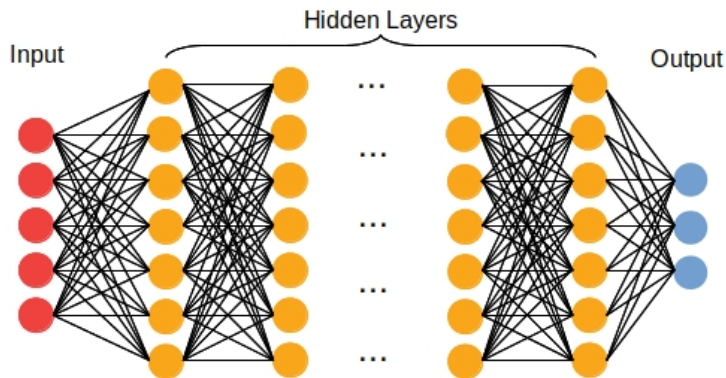


Figure 2.2: Architecture of deep neural networks.

The extracted features play a crucial role in the machine learning methods. Conventional machine learning methods achieve good performance by feature engineering, where multiple manually designed features are extracted to represent the given input image. Such features are typically known as hand-crafted features as these features are selected by experts based on the image characteristics and the domain knowledge. The hand-crafted features and associated parameters need careful design to fit different tasks. Prevalent features comprise the color features (e.g., color histograms), texture features (e.g., Local Binary Patterns (LBP) [3], Scale Invariant Feature Transformations (SIFT) [46], Histogram of Oriented Gradient (HOG) [17] and etc.) and morphological features (e.g., shapes). However, even with fine parameter setting and careful design, the extracted hand-crafted features usually fail in complex tasks as they are not representative and discriminative enough.

The DL method has become very popular and prominent in many computer vision applications. Especially, the CNN based methods provide encouraging results in the processing of images and videos. Typical architecture of CNN has been shown in Fig. 2.3. The current convolutional layers are obtained by

convoluting previous layers (input of current layers) with a set of 2-D convolution kernels. The pooling layers are obtained by pooling (usually subsampling by a factor of 2) operation given the previous layers as input. The CNN can automatically learn the optimal features and perform classification simultaneously with little or no preprocessing of the inputs. This is mainly because that the deep layer architecture can model the non-linear and complex relationship between the output and input by learning large number of parameters given sufficient training data.

While developing CAD systems for medical images has become a popular and promising direction, it is expensive and difficult to collect sufficient training data for medical image analysis. Transfer learning has been used, which aims at transferring knowledge learned in the source domain to the target domain, and improving the performance of the target learners [97]. One commonly used approach is using a pretrained neural network trained on a large dataset (e.g., ImageNet [18], an image database with more than fourteen million labeled images) as an initialization or fixed feature extractor. For example, Muazzam et al. finetuned AlexNet [36] to assist the classification and detection of Alzheimer’s disease based on MRI images. Similarly, Khan et al. finetuned VGG [73] for brain tumor classification with MRI images. In this PhD thesis, transfer learning has been used for the skin lesion and scattering image analysis. Because of the small size of scattering images, the transfer learning method that uses a pretrained neural network as an offline feature extractor has been used.

2.2 Automatic Skin Lesion Segmentation

A skin lesion has a higher chance of skin cancer if it has an irregular, blurred or jagged border. Hence, it is beneficial to first segment the ROI, which can benefit the subsequent skin cancer detection. Skin lesion segmentation can be a primary step for skin lesion classification and is also an important research area itself. Numerous segmentation methods have been developed for the skin lesion images. The pigment regions of suspicious lesions usually exhibit a

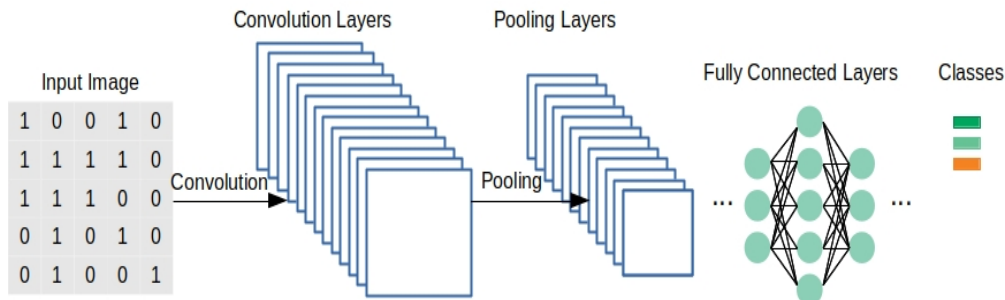


Figure 2.3: Typical architecture of CNN for a classification task. The current convolutional layers are obtained by convoluting previous layers with a set of 2-D convolution kernels. The pooling layers are obtained by subsampling the previous layers by a factor of 2.

darker color than the surrounding skin. Hence many conventional machine learning methods based on thresholding [32], region-merging [89], active contour models [1], [65] have been proposed. Among these methods, different clustering techniques with morphological operations have been used [4], [35]. Jafari *et al.* [35] used K-means clustering to segment the skin lesion into the foreground and background regions. Similarly, Ali *et al.* [4] used Fuzzy C-Means (FCM) to get the boundary of the pigment regions. Another popular class of approaches is the active contour models [1], [82], where the contour evolves toward the border of ROI in each iteration. [1] first get the candidate region by using the thresholding-based methods, active contour model driven by multi-direction Gradient Vector Flow (GVF) snake [82] or local histogram fitting energy [1] is then used to refine the coarse segmentation mask and make accurate segmentation. However, these conventional methods usually involve many intermediate steps (pre-processing and post-processing), which can affect performance of the proposed technique by a large margin. Such methods can achieve good performance with some simple datasets, where all the images share similar characteristics. Nevertheless, the real cases from clinics are more complex than the ideal situation. It will fail when the skin boundary is blurred and the skin conditions are complex. Moreover, the dermoscopic images collected using different instruments hold different specifications and characteristics, it is important to develop a technique that is optimized for datasets with different characteristics.

The CNN based methods have achieved great success in many computer vision applications, including skin lesion segmentation. The overall architecture of the segmentation network is similar to the classification network shown in Fig. 2.3, but usually with a decoder network. The neural network shown in Fig. 2.3 can be named as an encoder network, as it encodes the input image as a feature vector at the very end, before the fully connected layers. The decoder network usually shares a symmetric architecture with the encoder network but use upsampling in the pooling layers, which aims to increase the resolution of outputs and provide a segmentation mask with the same size as the input.

U-net [67] proposed by Ronneberger *et al.* uses a typical encoder and decoder architecture. It is very popular and has achieved huge success in medical images, and methods based on U-net [9], [14] for skin lesion segmentation and classification have been proposed. One can also perform the segmentation task and use the classification network with only the encoder architecture by performing pixel-level classification. Then, input is an image patch centered at a certain pixel, and the output is the probability of the pixel of being the foreground or background. The segmentation task is treated as a classification task in this case. Jafari *et al.* [34] proposed a skin lesion segmentation method using the Fully Convolutional Network (FCN). The inputs are image patches of different scales to make use of the local context information of multiple scales. However, the pixel-level classification requires dense prediction, which is time-consuming and computation-expensive. The global context information is not considered in this case. Hence, more recent works use the encoder and decoder architecture for the segmentation task [25], [90], [94]. Cross-entropy loss has been used as the standard loss function for optimizing the CNN models. Some methods have also been proposed, where novel loss functions are used to improve the performance of CNN. Yuan *et al.* [94] used Jaccard distance loss instead of the regular cross-entropy loss as the loss function and achieved better performance for skin lesion segmentation. The multi-task methods learn models by making predictions of multiple relevant tasks, given the extra labeling information. The extra labeling of these tasks can provide useful information which can promote the learning of the main task. Recently, [90] proposed

a mutual bootstrapping method based on CNN, where skin lesion segmentation and classification are performed simultaneously. Each task facilitates the other task in a bootstrapping way. More specifically, a coarse segmentation network is learned based on a standard CNN, which is then incorporated with the classification network to guide the learning of the classification network and make it focus on the ROI. At the same time, class-specific localization maps generated are concatenated into a segmentation network for the prediction of enhanced mask, which is more accurate. The multi-task framework requires datasets with extra labeling information, which is not easy to obtain for many medical images. State-of-the-art research has shown that combining the extreme points (points on the boundaries) with the original RGB images as the input of CNN can improve performance for segmentation tasks [51]. The extreme points provide extra information regarding the segment regions, and they can be obtained from the ground-truth segmentation masks without any extra labeling efforts. Based on this motivation, [25] proposed to use the RGB images and the corresponding predicted bounding boxes as inputs of a CNN model for skin lesion segmentation. The bounding box can be first learned by using Faster Region Based Convolutional Neural Networks (RCNN). Superior performance has been obtained compared with the methods that only use RGB images as inputs, and the best performance has been obtained for skin lesion segmentation.

2.3 Automatic Skin Lesion Classification

To detect skin cancers, multiple clinical metrics have been proposed using dermoscopy images based on the appearance of local patterns, including the Asymmetry, Border, Color, Diameter, Evolving (ABCDE) rules [28], [75], classical pattern analysis [61] and seven-point checklist [6]. The ABCDE rules provide an easy and quick way for the clinicians and patients to self-identification. The ABCDE rules are shorts for Asymmetry, Border irregularity, Color that is not uniform, Diameter greater than 6mm, and Evolving shapes (size, shape, or color). Fig. 2.4 presents some typical images of the normal and abnormal

(skin cancers) skin lesions that can be diagnosed with the ABCDE rules.





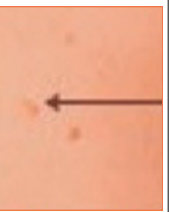



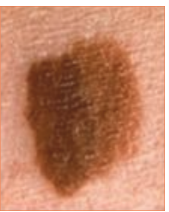
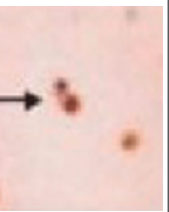
Criteria	A	B	C	D	E
Normal					
Abnormal					
Features	Half of the skin lesion look different from the other half in color and/or shape.	The skin lesion has an irregular, jagged or blurred border.	The skin lesion has different colors or different shades of one color.	Diameter of the skin lesion is larger than 6mm.	Change in size, shape or bump evolved.

Figure 2.4: Normal and abnormal skin lesions explained with the ABCDE rules. The abnormal skin lesion images are usually with asymmetric shape, irregular, jagged or blurred borders, multiple colors within the pigment regions, large diameter and evolving appearances over time.

However, the actual cases of suspicious skin lesions are far more complex. Different types of skin lesions can share a similar appearance, e.g., nevus and melanoma, which makes it even difficult for the experienced dermatologists to differentiate skin lesions accurately. Recent work has shown that the CNN based method can obtain comparable and even better performance compared with the experienced dermatologists [21], which demonstrates the potential and appealing prospect of automatic skin lesion analysis using the advanced CNN models.

As mentioned in section 1.1, the original skin lesion images suffer from lighting condition variation and interference from hairs and other artifacts. To tackle these problems, preprocessing techniques like hair removal [13], [64] and color enhancement [64] have been proposed for skin lesion analysis. After

the preprocessing step, color and texture features are extracted, and a classifier is then trained for skin lesion classification. The features play a key role in accurate prediction, and many conventional machine learning methods that utilize hand-crafted features have been proposed for skin lesion classification. Previous studies have demonstrated that the integration of various feature representations, such as color, texture, and shape features, is more powerful than a single type of feature representation [48], [59]. Unfortunately, hand-crafted features are not able to handle the complex tasks due to their inherent limitations, and they perform poorly in many cases for the skin lesion classification task.

Recently CNN based methods have achieved huge success in various research areas, and many popular CNN architectures have been proposed for the image classification tasks with good performance. One advantage of the CNN based methods is that it can automatically learn the optimal features and make predictions in an end-to-end manner. The features are task-specific and automatically learned during the training procedure based on the training data. Therefore, the deep features are semantically more representative and discriminative than the conventional hand-crafted features. The successful training of CNN depends on the initialization of the weights of neural networks. Many existing models for skin lesion classification use existing CNN models, and the parameters are optimized by finetuning with the skin lesion images as inputs. Model ensemble and test augmentation techniques are also used to further improve the experimental results [10], [50], [53], [54]. The multi-task framework has also been proposed for the skin lesion classification task [91], where the segmentation task and classification task are performed simultaneously. Generally speaking, the multi-task-based methods can obtain better performance than the single-task-based method since they can incorporate the shared information between different tasks. Existing CNN methods for the skin lesion analysis lack the interpretability of experimental results, which is important in assisting the clinicians in making decisions. [24] proposed a method named DermakNet, which used 50-layer ResNet [29] as the CNN backbone. Dermatologists' knowledge (e.g., attributes, asymmetry in-

formation) modeled by different sub-learning systems and meta-data are also used for interpretability and better performance. The CNN models can extract global semantic features of input images but ignore the local clues. To address this problem, [22] used both the global and local features for melanoma detection. The global features are extracted using standard ResNet [29] and the local features are obtained by using VGG-16 Network [73] with Bilinear Pooling (BP), which can distinguish skin lesions with subtle visual differences in local regions. To learn discriminative features and use images from different sources, [23] proposed a triple deep architecture, where a pair of digital and dermoscopic images of a single lesion is used as the input. Information from different sources is shared in the middle layers of neural networks. The features are then spatially weighted using the Class Activation Mapping (CAM) technique. BP is also used to generate the discriminative feature representation.

2.4 Scattering Image Analysis Based on Machine Learning

While dermoscopic images can be used for skin cancer detection, scattering patterns can also be used to detect the cancerous cells from healthy cells at the cell-level [41], [78], [81]. Laser scattering images provide essential information regarding cellular information and have been used in label-free cell identification. Conventional machine learning methods, as well as deep learning methods, have been proposed for scattering image analysis. Hand-crafted features such as the Gray Level Co-occurrence Matrix (GLCM), HOG have been used for scattering image analysis. The scattering images are dominated by randomly distributed white speckles on a dark background. [78], [79] showed that two speckle properties: the number and average area of speckles, are sufficient for accurate classification of different types of cells (CD34+ with mature myeloid precursor or Jurkat cells). [80] used GLCM features along with Support Vector Machine (SVM) to classify prostate cancer and healthy cells. [41] propose to use HOG features to discriminate ovarian cancer cells with the

normal cells. In contrast, deep learning methods can automatically learn the optimal features and classifiers in an end-to-end manner. A large number of parameters are used to learn the complex and non-linear relationships between the input and output by using multiple hidden layers. [45] proposed to use a simple CNN model for the classification of cervical cells using static cytometry. [81] used VGG [73], a popular well-known CNN architecture, to detect cancer cells in blood by multicolor stimulated Raman scattering.

2.5 Performance Evaluation Metrics

Different evaluation metrics are used for different tasks. Especially, to evaluate the segmentation results of skin lesions, the JA, Dice Coefficient (DC), Accuracy (ACC), Sensitivity (SEN) and Specificity (SPE) are used in this thesis. The criteria are calculated as follows:

$$JA = \frac{TP}{TP + FP + FN} \quad (2.1)$$

$$DC = \frac{2 \cdot TP}{2 \cdot TP + FP + FN} \quad (2.2)$$

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} \quad (2.3)$$

$$SEN = \frac{TP}{TP + FN} \quad (2.4)$$

$$SPE = \frac{TN}{TN + FP} \quad (2.5)$$

where TP, TN, FP and FN are the count of true positives, true negatives, false positives, and false negatives, respectively. JA represents the ratio of the overlapping area and the union area between the predicted segmentation mask and ground truth mask. DC is twice the overlapping area divided by the total number of pixels in both images. Both metrics reflect how close the prediction mask to the ground truth. SEN represents the proportion of foreground pixels being correctly segmented against the total number of foreground pixels. Similarly, SPE represents the proportion of background pixels being correctly segmented against the total number of background pixels.

To evaluate the skin lesion classification results, ACC and AUC have been used as the evaluation metrics for skin lesion classification, which is the main evaluation metric for skin lesion classification task and is widely used in the ISIC dataset [16].

For scattering image classification, ACC, SEN, SPE and AUC criteria have been used. The definitions of these criteria are introduced as follows:

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} \quad (2.6)$$

$$FPR = \frac{FP}{FP + TN} \quad (2.7)$$

$$TPR = \frac{TP}{TP + FN} \quad (2.8)$$

$$AUC = \int_0^1 T(F_0) dF_0 \quad (2.9)$$

where TP, TN, FP, and FN are the same definition as those in the segmentation task. False Positive Rate (FPR) and True Positive Rate (TPR) are shorts for false positive rate and true positive rate, respectively. $T(F_0)$ is the corresponding TPR when the FPR is F_0 . ACC measures the proportion of correct predictions (both true positives and true negatives) among the total number of cases. AUC provides an aggregate measure of performance against different classification thresholds.

Chapter 3

Skin Lesion Segmentation Using Deep Learning with Auxiliary Task

Skin lesion segmentation has been proved to be beneficial for the classification task and is an important research topic. For some skin lesions with light pigment, the colors and visual patterns of the pigment and the surrounding skin regions are very similar, resulting in fuzzy and unclear boundaries, making the skin lesion segmentation task extremely difficult. Moreover, the skin lesions also contain items such as hairs, veins, color-makers, rulers and glues, which affect the color and texture distribution of the skin lesions and impede successful learning. Fig. 3.1 displays some example images from the ISBI2017 dataset [16] for skin lesion analysis. To address these problems, literature studies that deploy different CNN architectures with multi-scale information [38], [44], [74], or multi-task learning framework [90], [91] have been proposed for skin lesion segmentation. The core idea of these methods can be regarded as trying to use as much information as possible to make robust predictions. However, these strategies either introduce extensive extra parameters for training or require extra labeling information, which may be inapplicable in practical situations.

In this chapter, a novel CNN method that uses auxiliary information is proposed for the skin lesion segmentation. The proposed method can be trained

The content of this chapter has been published as L. Liu, Y. Y. Tsui, and M. Mandal, "Skin lesion segmentation using deep learning with auxiliary task," *Journal of Imaging*, vol, 7, no. 4, 2021.

in an end-to-end manner without any pre-processing or post-processing steps. The contribution of this work is two-fold:

- Edge prediction is used in skin lesion segmentation as an auxiliary task. The proposed method learns the edge prediction and segmentation tasks simultaneously by two parallel branches. The edge prediction can guide the neural network to focus on the boundaries of the interest regions. Up to the authors' knowledge, this is the first work that uses edge information to help the skin lesion segmentation task. Note that the edge of a skin lesion can be automatically obtained by applying the contour detection technique and hence requires no extra labeling effort.
- A CCL module and a MSFA module have been proposed in this work. The CCL module can implicitly guide the learning, and boost each task's performance by feature interaction. The MSFA module can leverage the multi-scale information of the intermediate feature maps of different resolutions, whose weights can be automatically learned during training.
- Extensive experimental results verify that the edge prediction task can help the skin lesion segmentation task, and the proposed method achieves state-of-the-art performance.

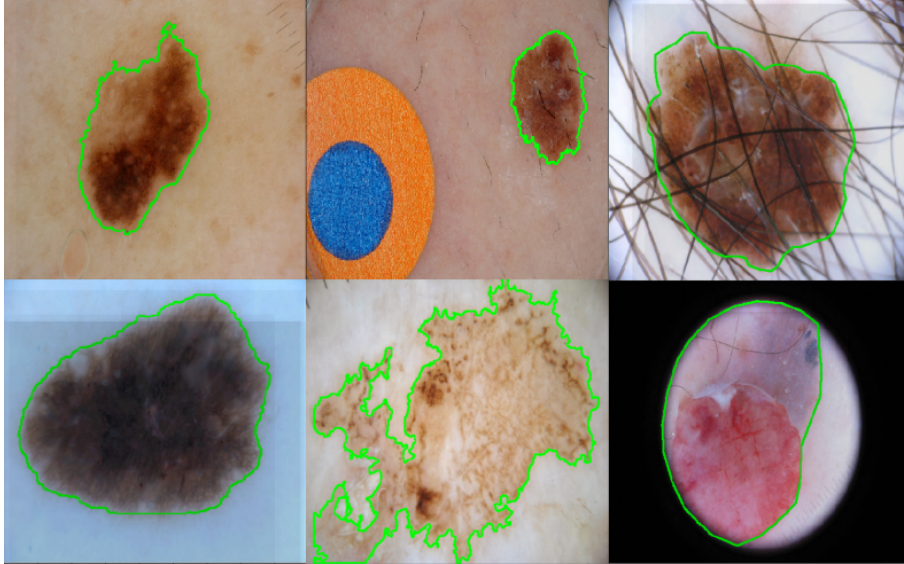


Figure 3.1: Some skin lesion images from the ISBI2017 dataset for skin lesion segmentation. The ground truth masks are marked using the green contours. Fuzzy boundaries and noisy items such as hairs and color-marks are also witnessed.

3.1 Proposed Method

In this section, we present details of the proposed method. During the training phase, the proposed method will simultaneously predict the segmentation mask and its corresponding edge of an input image. During the testing phase, only the segmentation mask is used for prediction. Schematic diagram of the proposed method is shown in Fig. 3.2. The proposed method consists of three main modules: the CNN backbone, the CCL module and the MSFA module. Details of each module are introduced in the following sections.

3.1.1 CNN Backbone

As shown in Fig. 3.2, an input image first goes through a backbone CNN to get the intermediate feature maps F for the edge prediction and mask prediction task. In this paper, ResNet-101 [29] and Pyramid Pooling Module (PPM) [96] are utilized as the backbone CNN, which can also be regarded as a strong baseline model for automatic skin lesion segmentation. The input images are first rescaled to 448×448 and then fed into the backbone CNN. To obtain

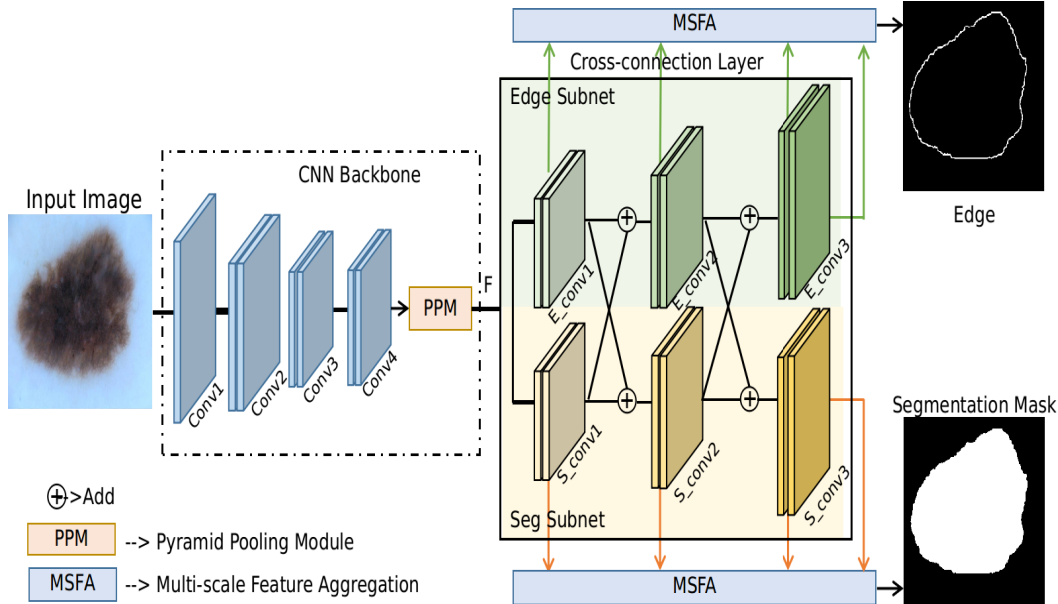


Figure 3.2: Schematic diagram of the proposed method. Edge prediction is used as an auxiliary task to assist the segmentation task. Two parallel networks with CCL module are implemented so that the learned two tasks can interact with each other during training and hence boost each other’s performance in turn. MSFA module is used to aggregate the multi-scale information from the intermediate feature maps of different scales.

a reasonable feature resolution of F , the Conv4 layer in the ResNet-101 is modified with stride and dilate rate has been set as 1 and 2, respectively [51]. By doing this, the resolution of the Conv4 layer will not decrease. Details about the backbone CNN architecture are shown in Table 3.1. Especially, $[1 \times 1, 64]$ represents that the filter size used during training is 1×1 , and the number of filters is 64, which also equals the number of output feature maps. $[\cdot] \times 3$ means the operations are applied three times sequentially. We set the stride to be 1 and use zero padding so that the size of the output feature maps will be the same. The PPM module [96] can leverage the context information of different resolutions and has been widely used as a plug and play tool. The same setting as [96] was used, except that the number of output channels was set to be 128. The PPM module fuses features under four different pyramid scales, with bin sizes of 1×1 , 2×2 , 3×3 and 6×6 , respectively. The input of the PPM module is the output feature maps of Conv4 layer with size $[1024, 56, 56]$, where 1024 is the number of feature channels and the spatial

Table 3.1: Architecture of the proposed method. The input image size is 448×448 .

layer name	output size	output channel dimension	operations
<i>Conv1</i>	224×224	64	$\begin{bmatrix} 7 \times 7, 64, \textit{stride} 2 \\ 3 \times 3, \textit{max pool}, \textit{stride} 2 \end{bmatrix}$
<i>Conv2</i>	112×112	256	$\begin{bmatrix} 1 \times 1, 64 \\ 3 \times 3, 64 \\ 1 \times 1, 256 \end{bmatrix} \times 3$
<i>Conv3</i>	56×56	512	$\begin{bmatrix} 1 \times 1, 128 \\ 3 \times 3, 128 \\ 1 \times 1, 512 \end{bmatrix} \times 4$
<i>Conv4</i>	56×56	1024	$\begin{bmatrix} 1 \times 1, 256 \\ 3 \times 3, 256 \\ 1 \times 1, 1024 \end{bmatrix} \times 23$
<i>F</i>	56×56	128	PPM [51]
<i>E_conv1/S_conv1</i>	112×112	128	$\begin{bmatrix} 1 \times 1, 128 \\ 3 \times 3, 32 \\ 1 \times 1, 128 \end{bmatrix} \times 1$
<i>E_conv2/S_conv2</i>	224×224	128	$\begin{bmatrix} 1 \times 1, 128 \\ 3 \times 3, 32 \\ 1 \times 1, 128 \end{bmatrix} \times 1$
<i>E_conv3/S_conv3</i>	448×448	128	$\begin{bmatrix} 1 \times 1, 128 \\ 3 \times 3, 32 \\ 1 \times 1, 128 \end{bmatrix} \times 1$

size of the feature maps is 56×56 . The output of the PPM module is feature maps of size $[128, 56, 56]$.

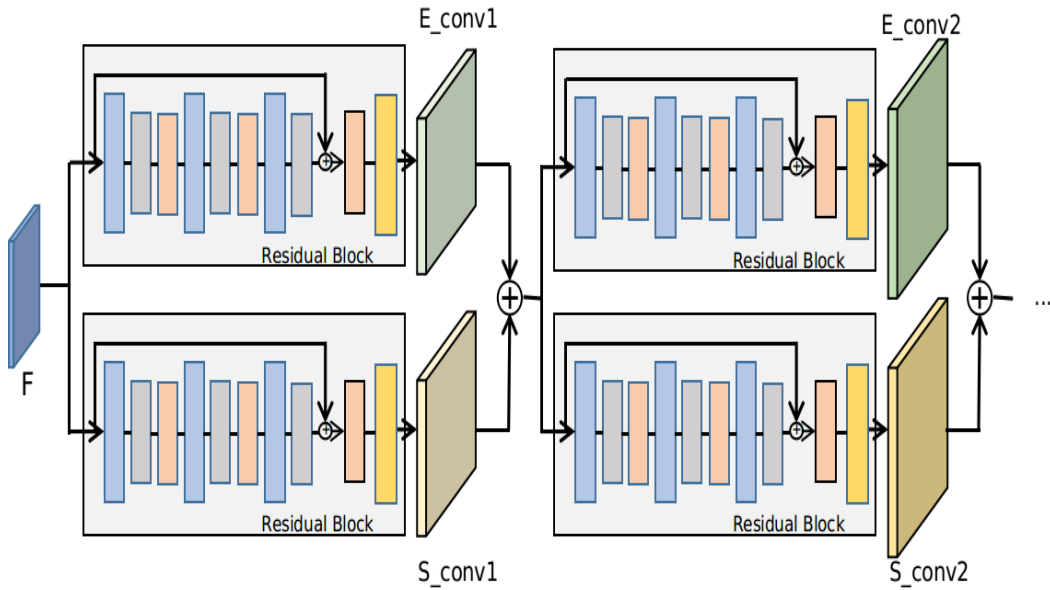
3.1.2 Cross-connection Layer (CCL)

The proposed method uses two parallel networks to predict the edge and segmentation mask simultaneously. Interactions between these two networks are realized by using the CCL module, where each task’s intermediate feature maps are used as the inputs of the next sub-block of the other task. In this case, the edge information is used during the forward prediction of the

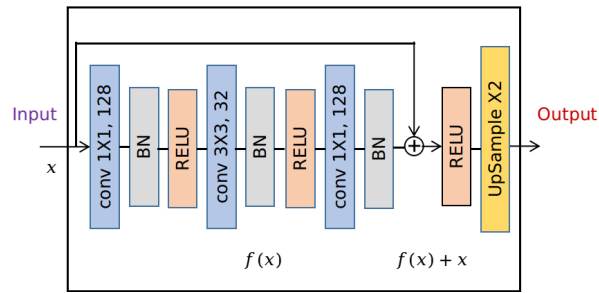
mask, which can implicitly regularize the boundary of the mask and guide the model to focus more on the contours. Similarly, the mask prediction subnet, which contains dense pixel and contextual information, is also used to guide the learning of the edges. $S_conv1, S_conv2, S_conv3$ are the sequential operation blocks of the Seg subnet, while $E_conv1, E_conv2, E_conv3$ are the sequential operation blocks of the Edge subnet. Specifically, the layers $S_conv1, S_conv2, S_conv3, E_conv1, E_conv2, E_conv3$ in CCL are obtained using the residual blocks, which have been shown in Fig. 3.3(a). Details about the implementation of the residual block are shown in Fig. 3.3(b). The residual block first uses 1×1 kernels to rescale the input feature maps. Afterward, 32 3×3 kernels are used, and hence the output channel number is shrunk to 32. The final output feature maps remain the same size as the input. This can be easily realized by using 1×1 kernels and setting the number of output channels to 128. Upsampling (by a factor of 2) is used after each residual block so as to increase the feature resolution. The size of the output feature maps at layer S_conv3 and E_conv3 is 448×448 , which is equal to the size of the input images.

3.1.3 Multi-scale Feature Aggregation (MSFA)

The MSFA module, shown in Fig. 3.4(a), is used to aggregate the feature maps from each resolution and make the final prediction. It first uses the *Conv* block to generate an output prediction map for feature maps at each resolution, which can be regarded as the process of making predictions at a certain scale. The *Conv* block consists of standard convolution operations: first, convolution with 3×3 kernels (the number of channels is 128) is performed. Batch Normalization (BN) and Rectified Linear Unit (RELU) are used afterward. The output of the *Conv* is a feature map with depth one, which is obtained by convolving with a 3×3 kernel. For instance, feature maps at layer $S_conv1, S_conv2, S_conv3$ are fed into the MSFA module and three feature maps are generated at different scales via the *Conv* block. To leverage the feature maps at higher resolution more effectively, we also generate the fourth feature map by directly convoluting S_conv3 with a 1×1 kernel. These four

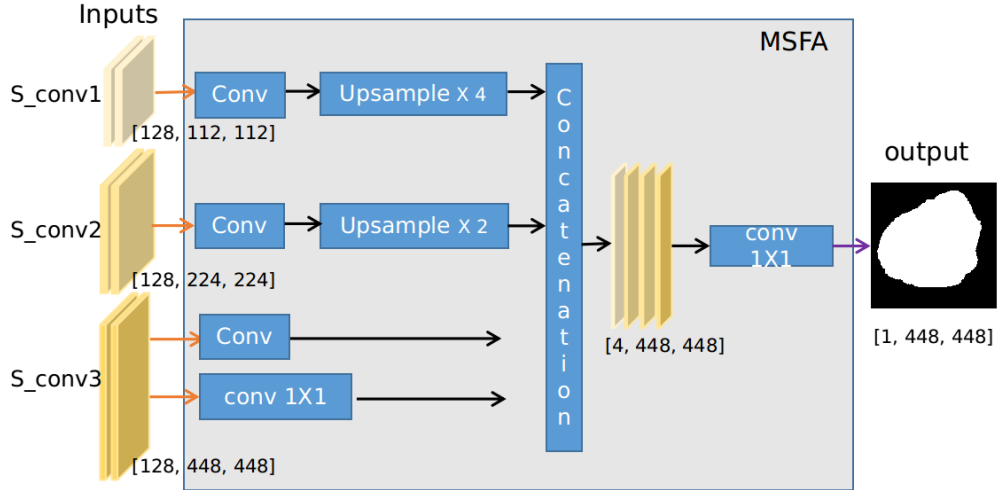


(a) CCL (Cross-connection Layer) module

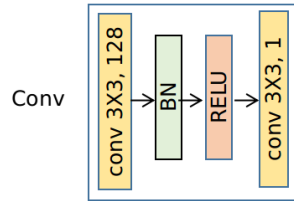


(b) Implementations of the Residual Block

Figure 3.3: Implementations of the CCL module. The inputs are feature maps F obtained from the backbone CNN. Fig. 3.3(a) shows the implementation details of a Residual block in Fig. 3.3(a), and blocks with the same color indicate the same operation.



(a) MSFA (Multi-scale feature aggregation) module



(b) Conv block of MSFA

Figure 3.4: Implementations of the MSFA module, where the segmentation branch is used as an example. The inputs are feature maps of different scales from the segmentation branch, e.g., S_{conv1} , S_{conv2} , S_{conv3} . Four prediction masks are obtained after some convolution and pooling operation. The final output is weighted sum of these predictions which can be automatically learned by 1×1 convolution.

predictions are then upsampled to size 448×448 and are concatenated, resulting in feature maps of size $4 \times 448 \times 448$. Finally, these four prediction masks are convolved with 1×1 kernel to aggregate these feature maps into one final prediction. The weight for the prediction of each scale can be automatically learned in this case. Fig. 3.4 shows the schematic of the MSFA module for the Seg subnet, the schematic of the MSFA module for the edge subnet is similar. For both segmentation and edge prediction, shared parameters are used in the MSFA module, which encourages the prediction masks and edges to share a similar quality.

The class-balanced cross-entropy loss is used as the cost function for the segmentation and the edge prediction task. The class-balanced cross-entropy

loss l of a prediction is calculated using the following equation:

$$l = -\frac{1}{N} \sum_{n=1}^N [w_1 \times y_n \times \log(h_\theta(x_n)) + w_0 \times (1 - y_n) \times \log(1 - h_\theta(x_n))] \quad (3.1)$$

where

N : number of pixels;

y_n : target label for pixel n ;

x_n : input pixel n ;

h_θ : model with neural network weights θ ;

w_1 : weight for foreground pixels;

w_0 : weight for background pixels;

w_1 and w_0 are the class balanced weights and can be calculated by inverse class frequency: $w_1 = \frac{N_{neg}}{N}$ and $w_0 = \frac{N_{pos}}{N}$. N_{neg} and N_{pos} represent the number of background and foreground pixels of a ground truth mask, respectively. The losses of the segmentation and edge prediction functions can be very different, hence, a parameter α is used to balance these two tasks. The final loss function L is given by:

$$L = \alpha L_{seg} + L_{edge} \quad (3.2)$$

where L_{seg} and L_{edge} are the loss for segmentation and edge prediction task over the entire training data, respectively. Both L_{seg} and L_{edge} use Eq. (3.1) to calculate the loss. For the proposed method, we manually set $\alpha = 0.05$ to make the L_{seg} and L_{edge} in a similar range of values during training. The proposed method can be trained in an end-to-end manner.

3.2 Dataset

For performance evaluation of the proposed method, we have used the dataset from ISIC 2017 for skin lesion detection [16], which is a very challenging dataset for skin lesion classification. There are 2000 images in the training set, including 374 melanoma, 254 seborrheic keratosis, and 1372 benign nevi. The

validation dataset contains 150 images and the final testing dataset contains 600 images. All the images are of various resolutions, ranging from 767×1022 to 4499×6748 pixels. Severe illumination variation, noise and various artifacts are also witnessed in this dataset.

3.3 Experimental Results

In this section, we first present implementation details of the proposed approach. Ablation studies have been conducted to show the benefits of each module. Finally, the technique’s performance is evaluated and compared with the state-of-the-art techniques.

3.3.1 Implementation Details

The proposed model is learned using the training data of ISBI2017, and the performance is evaluated on the testing data. To train the proposed model, a dermoscopic image (input), its corresponding ground-truth segmentation mask and edge (contour) image (outputs) are required. The ground truth of the edge image can be automatically obtained from the ground truth of the segmentation mask by contour detection technique. During the training phase, data augmentation is used to increase the number of training images. Data augmentation techniques, including random horizontal and vertical flipping, center cropping at a random scale between $[0.75, 1.25]$, random rotation in the degree range $[-20, 20]$, ground truth cropping with zero-padding of 50 pixels, and an image deformation method named Rigid Moving Least Squares [70] are used to generate more training images. All the images are then rescaled to the size of 448×448 . To train the proposed neural network, we set the batch size to be 8 and train it for 30 epochs. Adam optimization algorithm with an initial learning rate of 0.0001 is used, which decreases exponentially with a learning rate decay $\gamma = 0.9$.

During the testing phase, only the segmentation mask is needed to evaluate the performance of the proposed method. Test augmentation is also performed by rotating the input test image by 90° , 180° , 270° , flipping horizontally and

vertically. The prediction of the original image can be obtained by reverse operation of the predictions to the transformed images. The average of these prediction outputs is then used as the final segmentation output.

The skin lesion segmentation method is implemented using PyTorch. All the experiments were conducted on a server with an Intel Xeon Processor (Skylake, IBRS) CPU and two GPUs of Nvidia Tesla V100 with 16 GB memory.

3.3.2 Parameter Setting of the Loss Function

As in Eq. 3.2, there is one important parameter α that balances the loss between the segmentation task and the edge prediction task. To investigate the impact of parameter α , experiments regarding different values of α are conducted. The experimental results are shown in Fig. 3.5, where JA is used as the evaluation metric. More results are provided in Table 3.2.

As we can see from Fig. 3.5 and Table 3.2, $\alpha = 0.05$ can give us the best performance regarding the JA. Performances of the proposed method are relatively stable when the values of α are smaller than 0.25, and the JA gradually decreases with increasing values of α . For instance, the JA drops from 79.46 to 79.01 when the value of α increases from 0.05 to 0.5. This is because a larger value of α suggests a weaker role of edge prediction, which has proved that the edge prediction task can benefit the segmentation task. The proposed model will focus more on the boundaries of ROI, which is crucial for successful segmentation. Notice that in the extreme case, only the segmentation branch is updated and learned during the training phase when the parameter α is significantly large. Therefore, in this study, we have set $\alpha = 0.05$ to obtain a balanced segmentation and edge prediction loss.

Table 3.2: Experimental results with different values of α on the test set.

α	ACC	DC	SEN	SP	JA
0.005	94.17	87.14	88.77	95.56	79.43
0.05	94.32	87.13	88.76	96.51	79.46
0.25	94.33	87.09	88.06	96.40	79.30
0.5	94.11	86.78	89.25	93.39	79.01

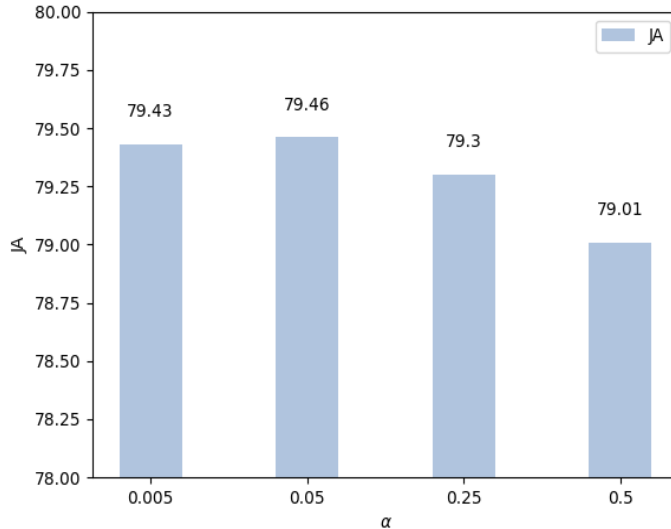


Figure 3.5: Experiment results with different values of α on the test set.

3.3.3 Ablation Study

To show the effectiveness of the proposed method, an ablation study of the proposed method on the ISBI2017 test data is conducted. We name the three yellow blocks (in Fig. 3.2) in the segmentation branch as the Seg Subnet. Similarly, the three green blocks in the edge prediction branch are named the Edge Subnet. Results of the ablation study are shown in Table 6.1. ResNet+PPM+Seg is the backbone CNN with a Seg subnet for the segmentation task, which can be regarded as a strong baseline model for the skin lesion segmentation task. JA is regarded as the main evaluation metric for the segmentation task as in the literary works, which reflects the percentage of overlap between the prediction mask and the ground-truth mask. A JA value of 77.01% is obtained for this baseline method. By adding the Edge subnet, we further increase the JA by 0.57%, which verifies that the auxiliary task (edge prediction) can benefit the segmentation mask. Our final model is the proposed method with the Seg subnet, the Edge subnet, and the MSFA module, which obtains the best performance with JA value of 79.46. An increase of 2.45% JA value is observed for the proposed method compared with the baseline method, which verifies the effectiveness of the proposed method.

In addition to JA, the proposed method provides an improvement of 1.16%, 1.92%, and 1.39% for ACC, DC and SPE over the baseline method.

Table 3.3: Ablation study of the proposed method .

Method	ACC	DC	SEN	SPE	JA
ResNet+PPM+Seg	93.16	85.21	88.87	95.12	77.01
ResNet+PPM+Seg+Edge	93.54	85.66	87.11	96.61	77.58
Proposed	94.32	87.13	88.76	96.51	79.46

3.3.4 Comparison with State-of-the-art Methods

In this section, we compare the proposed method with other state-of-the-art methods using the ISBI2017 test data. The experimental results are shown in Table 6.2. Since ISBI2017 is a challenge dataset for skin lesion segmentation, ensembling techniques by using different CNN models [93], post-processing [11], [43], [93] are widely used. In comparison, our method only learns one model in an end-to-end manner without any pre-processing and post-processing methods and still achieves the best performance with JA of 79.46, ACC of 94.32, SEN of 88.76.

Table 3.4: Experimental results compared with state-of-the-art methods on ISBI2017 test data.

Method	ACC	DC	SEN	SEN	JA
Liu <i>et al.</i> [43]	93.00	84.00	82.90	98.00	75.20
Abhishek <i>et al.</i> [2]	92.22	83.86	87.06	95.16	75.70
Yuan <i>et al.</i> [93]	93.40	84.90	82.50	97.50	76.50
AI-Masni <i>et al.</i> [52]	94.03	87.08	85.40	96.69	77.11
Bi <i>et al.</i> [11]	94.08	85.66	86.20	96.71	77.73
Sarker <i>et al.</i> [69]	93.60	87.80	81.60	98.30	78.20
Proposed	94.32	87.13	88.76	96.51	79.46

3.4 Discussions

3.4.1 Quantitative Analysis of Test Results

In this section, quantitative analysis about the performance of the proposed method on the test data is also conducted. The histograms regarding the JA

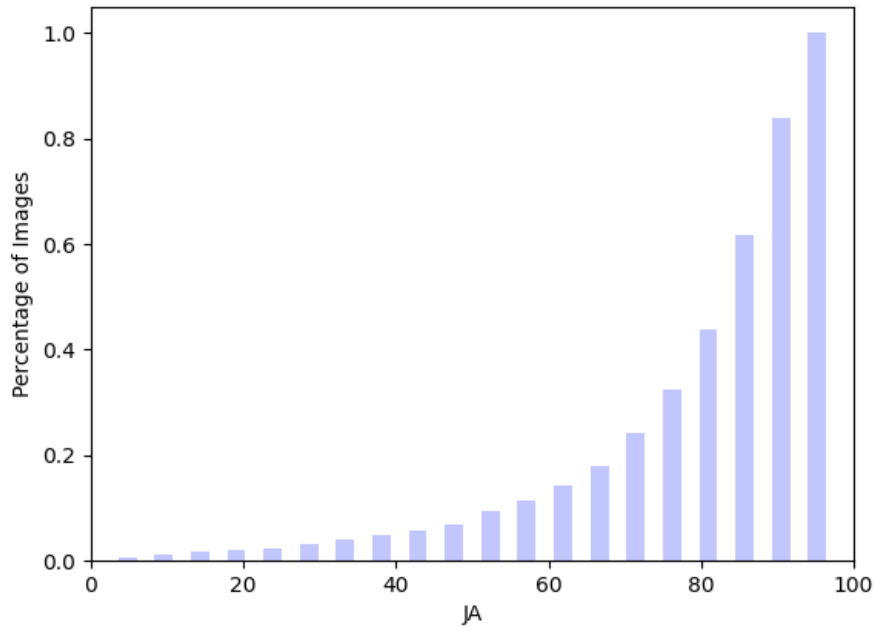


Figure 3.6: Cumulative histograms of the JA values on the test data.

distribution on the test data are shown in Fig. 3.6. As shown in Fig. 3.6, over 70% of test images can be segmented with JA larger than 70.

3.4.2 Qualitative Analysis of Test Results

The final outputs of the edge prediction and segmentation branches are displayed in Fig. 3.7. As shown in Fig. 3.7, the proposed approach can segment the ROI accurately in most cases. The first row displays the predictions given an easy input image, whose color contrast is high between the foreground and background regions. The proposed approach can detect the pigment region with high accuracy. The second, third, and fourth rows present the predictions of input images with fuzzy boundaries and low contrast. In addition, the existence of glue is also observed among these images, which will make it extremely difficult to identify the boundaries. Output prediction maps of the edge prediction and segmentation branches become slightly fuzzy on the boundaries in these cases, but still with decent results. The bottom row shows the predictions given an input image with hairs and low color contrast, which

can affect the textures of the skin lesions and prohibit successful learning. Nevertheless, the proposed method still successfully segments the pigment regions. In other words, the proposed method is robust to noisy items and obtains an overall good performance.

Fig. 3.7 shows some test images that have been segmented successfully. To further evaluate the performance of the proposed approach, we have also display some failed cases in Fig. 3.8. The first row displays the input images, the second row shows the corresponding ground truth masks, and the third row is the prediction mask of the proposed approach. As we can see from the first two columns in Fig. 3.8, the proposed approach generates larger prediction masks than the ground truths, which is mainly due to the existence of the glue and the unclear boundaries. They will distract the proposed method and deteriorate the segmentation performance. In contrast, outputs in the third and fourth columns show that the proposed method predicts smaller segmentation masks than the ground truth masks. The learned model tends to treat the dark areas as the foreground regions (which is the most frequent cases for dermoscopic images), and it fails when there are multiple colors scattered within the same lesions in some difficult cases. For instance, the proposed method fails when the foreground region contains a dark area surrounded by an area of light color, whose appearance is more similar to the healthy skin region (see the last two columns of Fig. 3.8). Such a phenomenon has also been found in previous works [84], [88], which may be caused by the scarce samples and it will be the focus of our future research. It is also worth noting that the ground truths of the segmentation masks of ISBI2017 are labeled using different methods (e.g., manual labeling, thresholding methods, interactive labeling methods). Therefore, there are annotation disagreements among the labeled images, which has been described in [25], [66] and may hinder the learning of the proposed method.

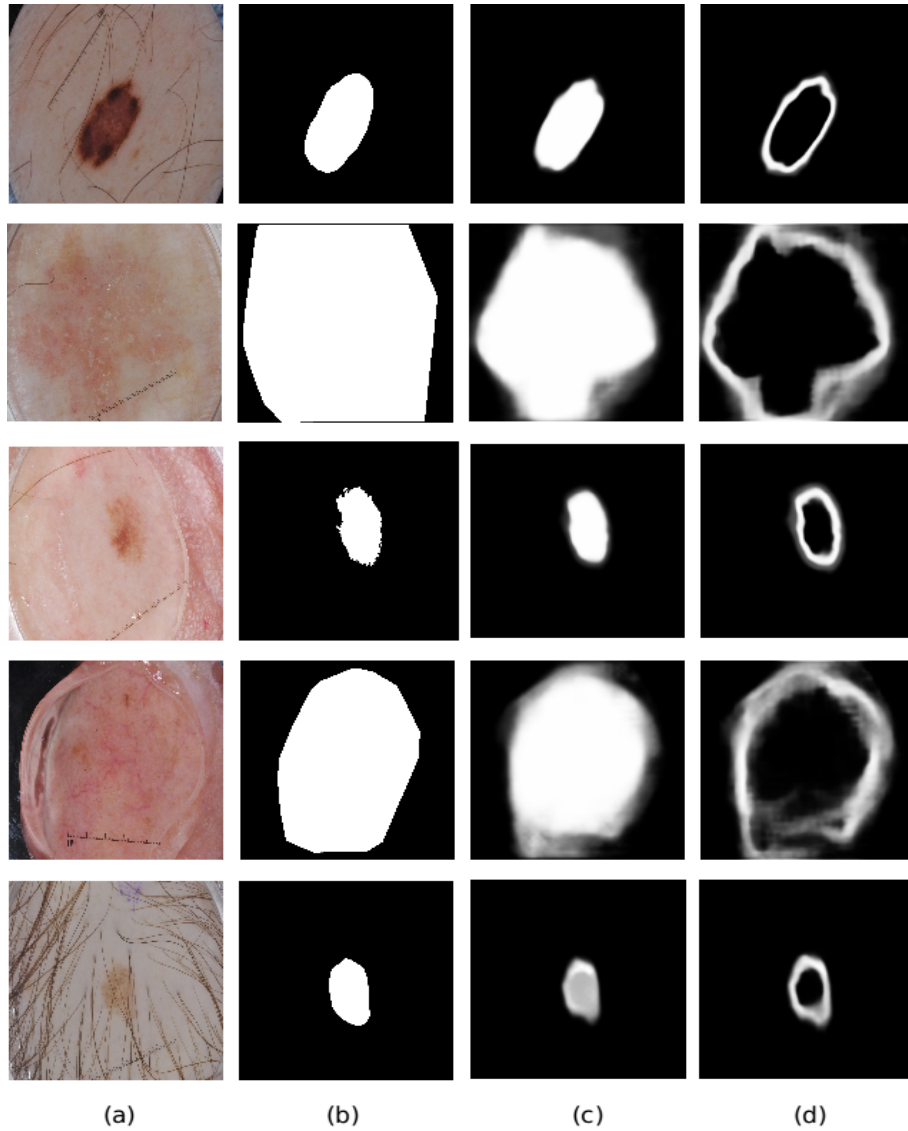


Figure 3.7: Output visualization of the proposed method. (a) input test image; (b) the corresponding ground truth segmentation mask; (c) the output probability map of the segmentation prediction branch; (d) the output probability map of the edge prediction branch.

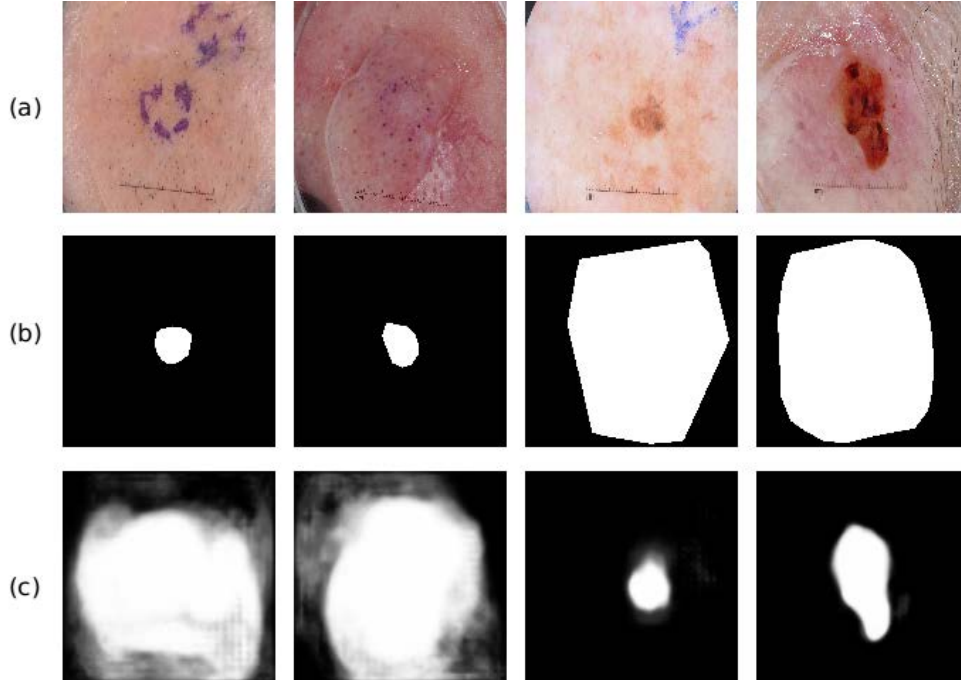


Figure 3.8: Some examples about the failed cases for the proposed method. (a) the input test images; (b) the corresponding ground truth segmentation masks; (c) the probability map of the proposed method.

3.5 Conclusions

In this chapter, a novel CNN based method with auxiliary task learning is proposed. Edge prediction, as an auxiliary task, is performed simultaneously with the segmentation prediction to help the segmentation task. The ground truth of the edge prediction task can be obtained automatically from the ground truth segmentation masks by using a standard contour detection method, and hence no extra labeling effort is required. A CCL module is proposed, where the intermediate feature maps of each task are fed into the other task’s sub-block, which implicitly guides the neural networks to focus on the boundary region and boosts the performance of the segmentation task. A MSFA module is proposed, which can automatically learn the final mask by aggregating the output of different scales. An ablation study has shown the benefits of these proposed modules. Experimental results with the ISBI2017 dataset have shown that the proposed method outperforms the other state-of-the-art methods in terms of performance measures such as the JA and accuracy.

Chapter 4

Automatic Skin Lesion Classification Based on Mid-level Feature Learning

In this chapter, a novel mid-level feature learning method is proposed for the skin lesion classification. Our motivation is that: the dermoscopic images suffer from strong visual similarities among different types of skin lesions and visual variations within the same class of samples. Several image pairs with their Euclidean distances values d are displayed in Fig. 4.1. The Euclidean distance values are calculated by using the features extracted via pretrained ResNet [29]. The average distance for intra-class pairs of images is 37.5, while for some hard samples, the intra-class distances can be larger than the average value. Similarly, the inter-class distances can be smaller than 37.5. Therefore, it is very difficult to learn an optimal feature representation that can well separate all the training images. Instead of using the original features as the input, relationships among different sample images are used as the feature representation. The relationships are modeled using the similarity measurement based on metric learning (learned using the training set) with a given reference set. An SVM classifier is finally used for the classification task by using the mid-level features as input.

The content of this chapter has been published as L. Liu, L. Mou, X. Zhu, and M. Mandal, "Automatic skin lesion classification based on mid-level feature learning," *Computerized Medical Imaging and Graphics*, vol. 84, 2020.

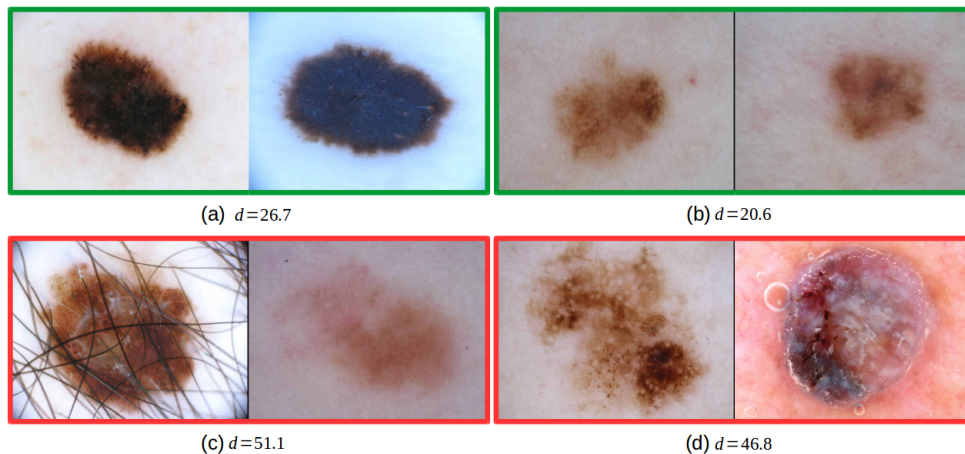


Figure 4.1: Example images from ISIC 2017 dataset on skin lesion analysis towards melanoma detection. Each green box indicates a pair of inter-class samples (left: melanoma vs. right: benign), and each red box indicates a pair of intra-class samples (both images are melanoma). Strong inter-class visual similarity and intra-class variations are observed across different types of skin lesions.

We name it mid-level feature representation since it captures higher-level affinity information of the original features, and can be regarded as an intermediate semantic feature representation, which bridges the raw features and the classification task. The affinity information of an image pair can be obtained by calculating their distance values. However, the Euclidean distance, although widely used in machine learning, cannot capture precise relationships among different image pairs, due to its limitations and incapacities. Therefore, Mahalanobis distance is learned by using a distance metric learning method. The distance metric learning can be regarded as a method for learning discriminative feature representation. Compared with the discriminative features learned by metric learning, the proposed mid-level feature is a soft discriminative feature representation, where the relationships of visual similarities and distinctions can be kept for some difficult cases (hard samples) as long as the remaining relationships are captured correctly. The learned image features are thus more robust to the large visual similarities between different classes of skin lesions and noisy items.

Specifically, a CNN based skin lesion segmentation model is used as a primary step for ROI detection. The skin lesion segmentation method is based

on U-net [67] and dilated convolution [92]. Features are then extracted from the ROI via the pretrained neural networks (ResNet [29] and DenseNet [31]). Finally, SVM is used to perform skin lesion classification. Block diagram of the proposed approach has been shown in Fig. 4.2.

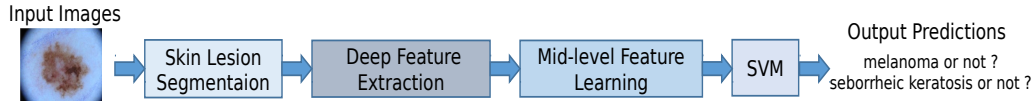


Figure 4.2: Block diagram of the proposed method.

The contributions of this work can be summarized as follows:

- A novel mid-level feature representation that utilizes the relationships among image samples (e.g. between an input image and the reference image set) is proposed for skin lesion classification. The new feature representation contains high-level affinity information between samples. It is a soft discriminative feature, having more tolerance to difficult cases, and is more robust to noise, large inter-class similarity and intra-class variations.
- A novel framework that integrates skin lesion segmentation has been proposed for skin lesion classification. Pretrained CNN models are used as the off-the-shelf feature extractors of ROI, and metric learning is then utilized to learn the mid-level features for classification.
- Extensive experiments have been conducted to show the advantages of the proposed approach, and experimental results show that the proposed method outperforms state-of-the-art CNN based methods.

4.1 Proposed Method

The proposed method contains four steps: skin lesion segmentation, feature extraction, mid-level feature learning, and SVM classification. Details about each step are presented in the following sections.

4.1.1 Skin Lesion Segmentation with Improved U-Net

Skin lesion segmentation is still a challenging task due to the complex skin conditions and blurred boundaries. To address the aforementioned problems, an improved CNN model based on U-net architecture is proposed in this section. Different from the original U-net, batch normalization layer [33] and dilated convolution [92] are introduced in this method. The batch normalization layer can tackle the internal covariate shift problem [33] and promote the performance of CNN. Dilated convolution is used to enlarge the perceptive field without losing resolution during the training procedure. Moreover, a simple test augmentation technique is proposed in this paper, which boosts the performance significantly. Compared with existing ensemble techniques, which usually need to train multiple models, our method integrates results only in the testing phase, which does not require extra training efforts, thus is very efficient.

Schematic plot of the proposed method model is shown in Fig. 4.3. The input image consists of 6 channels corresponding to RGB and HSV color spaces, and the output is the (segmentation) probability map. The number of feature maps are shown under each operation block. All convolutional kernels used in this model are of size 3×3 . As we can see from Fig. 4.3, the proposed method contains an encoder network and a decoder network. The encoder blocks use the following structure: *conv*, *BN*, *conv*, *BN*, *dilated_conv* and *max_pool*. The *conv* is the convolution layer, *BN* is the batch normalization layer, *dilated_conv* is the dilated convolution layer at stride of [2, 2], and *max_pool* is the max pooling layer with factor 2. For each decoder block, upsampling layer of factor 2 is first used to increase the feature maps' resolution. We then concatenate these feature maps with feature maps of the same size from the encoder path (as the green arrows show in Fig. 4.3). The overall structure of the decoder block is: *up_sampling*, *concatenate*, *conv*, *BN*, *conv*, *BN* and *dilated_conv*. The yellow block Conv_5 in Fig. 4.3 is the connection layer for the encoder network and

The content of this section has been published as L. Liu, L. Mou, X. Zhu, and M. Mandal, "Skin Lesion Segmentation Based on Improved U-net," *IEEE Canadian Conference of Electrical and Computer Engineering*, 2019, pp. 1-4.

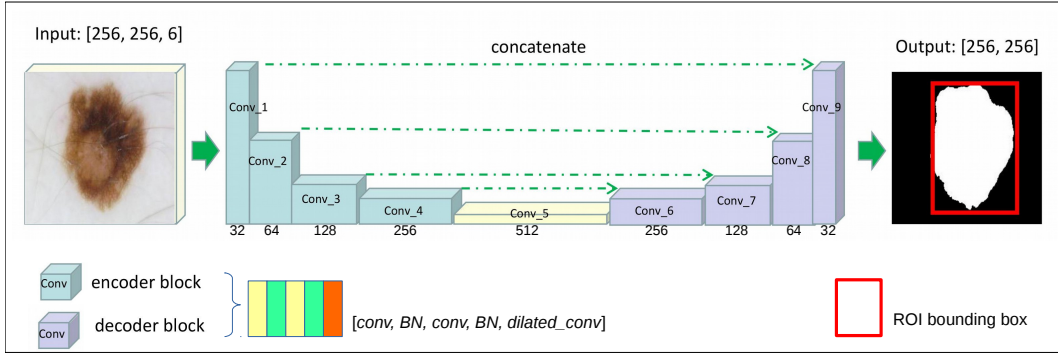


Figure 4.3: Schematic of the proposed CNN based model for skin lesion segmentation. The proposed method contains an encoder network and decoder network, each network is consisted by a sequence of encoder blocks and decoder blocks.

decoder network, and it is consisted of *conv*, *BN*, *conv*, *BN* and *dilated_conv*.

The original images are of high resolution, to train the deep model more efficiently, we first resize the input image size 256×256 pixels. RGB and HSV color channels (i.e., 6 channels) are used as the inputs. Each color channel is normalized to the range of $[0, 1]$ by using standard min-max normalization. During the testing phase, a simple test ensemble technique is used by data augmentation with the test data. For a test image, we obtain its augmented images by rotating the original images by 90, 180, 270 degrees, flipping them vertically and horizontally. Predicted masks of these transformed test images can be obtained using the trained model. We can get the predictions of the original image by reverse operation by using the transformed masks. The final prediction mask is the average values of these reconstruction maps and the original prediction mask.

For all the convolution layers, the stride is set to be 1, same padding and Rectified Linear Units (ReLU) are used. Adam optimization algorithm is used for neural network training. The learning rate is set to be 0.0001, and cross-entropy loss is used as the lost function. The number of epochs during training is set to be 3, and the batch size is set to be 16. Therefore, the steps per epoch equal to $48000/16 = 3000$.

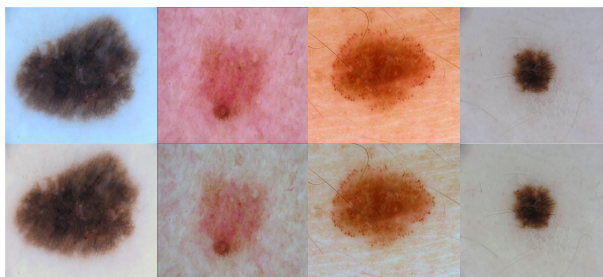


Figure 4.4: The first row shows the original images, and the second row shows the images after pre-processing using Retinex method. The color distributions of different images are enhanced via the use of Retinex method. The resulted images are of similar lighting conditions.

4.1.2 Deep Feature Extraction

In this section, deep features are extracted from the ROI images using the pretrained neural networks. The pretrained neural networks can extract rich and meaningful texture information of images, and have been successfully used as offline feature extractors for medical image analysis [27], [50]. In this study, pre-trained ResNet-50 [29] and DenseNet-201 [31] are used as off-the-shelf feature extractors. We denote them as ResNet and DenseNet for clarity in the following paper. Before feature extraction, Retinex algorithm [20] is used to enhance the color consistency among different images. The resulting images are shown in Figure 4.4. The colors of different images are more comprehensive and consistent after using the Retinex method. The ROI bounding boxes obtained from the previous segmentation module are superimposed on the pre-processed images to obtain the ROI images, which are then resized to 224×224 for feature extraction. For both CNN, the output of the Global Average Pooling (GAP) layer is used as the feature. The output features corresponding to the two CNN are of dimension 2048 (ResNet) and 1920 (DenseNet), respectively. Principal Component Analysis (PCA) is used to reduce the feature dimension by keeping 99% energy. The reduced feature dimensions for ResNet and DenseNet are 700 and 532, respectively.

4.1.3 Mid-level Feature Learning

Due to the complex skin conditions, noise, artifacts, and severe visual similarities among different types of skin lesions, the extracted features may have limitations in describing characteristics of the original data and have poor discrimination power. Instead of using the original features as input, a novel mid-level feature representation, which describes relationships among images, is learned and used as the input of the SVM classifier. The mid-level features of an image are obtained by learning the similarities between a given image and a reference image set. Since the discriminative power of the original features on the Euclidean space is poor due to the strong visual similarities among different classes of skin lesions, metric learning is used to address this problem. The metric learning method can learn a similarity measure to separate samples of different classes. Here, we present a brief introduction to the metric learning method. The squared Euclidean distance between two features x_i and x_j can be calculated by:

$$d(x_i, x_j) = (x_i - x_j)^T(x_i - x_j) = (x_i - x_j)^T I(x_i - x_j)$$

where I is an identity matrix. Similar to the formulation of Euclidean distance, instead of using an identity matrix, the Mahalanobis distance between features is defined as:

$$d_M(x_i, x_j) = (x_i - x_j)^T M(x_i - x_j), \quad s.t. \quad M \geq 0 \quad (4.1)$$

where M is a positive semidefinite matrix to be learned during the training procedure. Since M is a positive semidefinite matrix, it can be represented as $L^T L$, and the above function can be reformulated as:

$$\begin{aligned} d_M(x_i, x_j) &= (x_i - x_j)^T M(x_i - x_j) \\ &= (x_i - x_j)^T L^T L(x_i - x_j) \\ &= \|L(x_i - x_j)\|^2 \\ s.t. \quad &M \geq 0 \end{aligned} \quad (4.2)$$

By observing Eq. 4.2, the distance metric learning method can also be treated as a discriminative subspace learning problem that aims at learning L , and the new discriminative feature of x_i is denoted as Lx_i . The metric learning method

expects the distance between within-class samples to be small, and the distance between inter-class samples to be large. Compared with the discriminative feature Lx_i , the learned feature in this paper is a soft discriminative feature. Experiments in Sec. 4.4.4 have shown the advantage of the soft discriminative features over the discriminative features.

The optimal distance metric can be learned by separating images of the same class and different classes by a margin of μ . The objective with the logistic loss is shown in the following:

$$f_M(x_i, x_j) = \log(1 + e^{y_{ij}(d_M(x_i, x_j) - \mu)}), \quad (4.3)$$

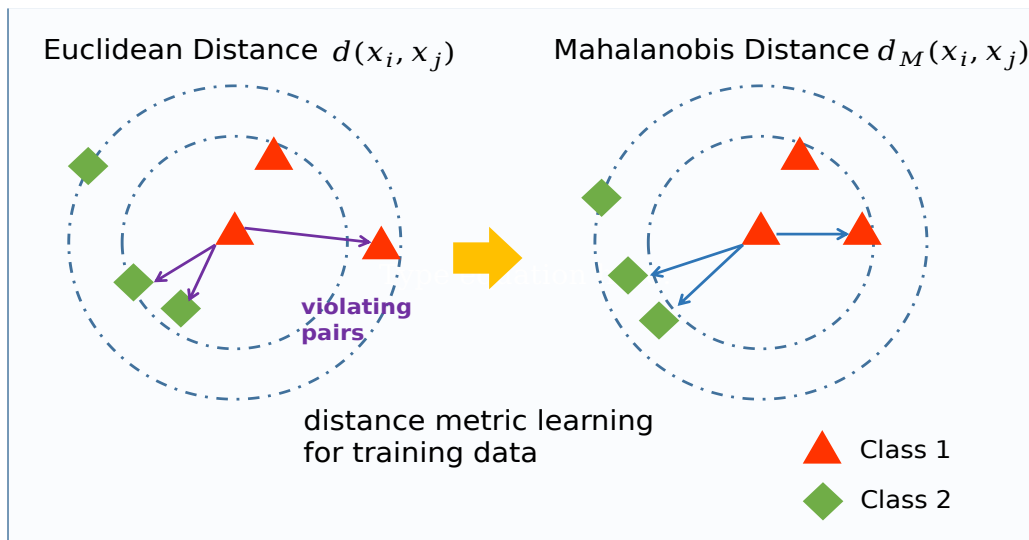
$$y_{ij} = \begin{cases} 1 & \text{if } y(x_i) = y(x_j) \\ -1 & \text{if } y(x_i) \neq y(x_j) \end{cases}$$

where $y(x_i)$ is the label of input feature x_i . The above function can drive distances between intra-class instances to become smaller than μ , and distances between inter-class instances to become larger than μ . In this paper, we set μ as the average Euclidean distance between samples of the same class. The optimal solution is learned by minimizing the following function:

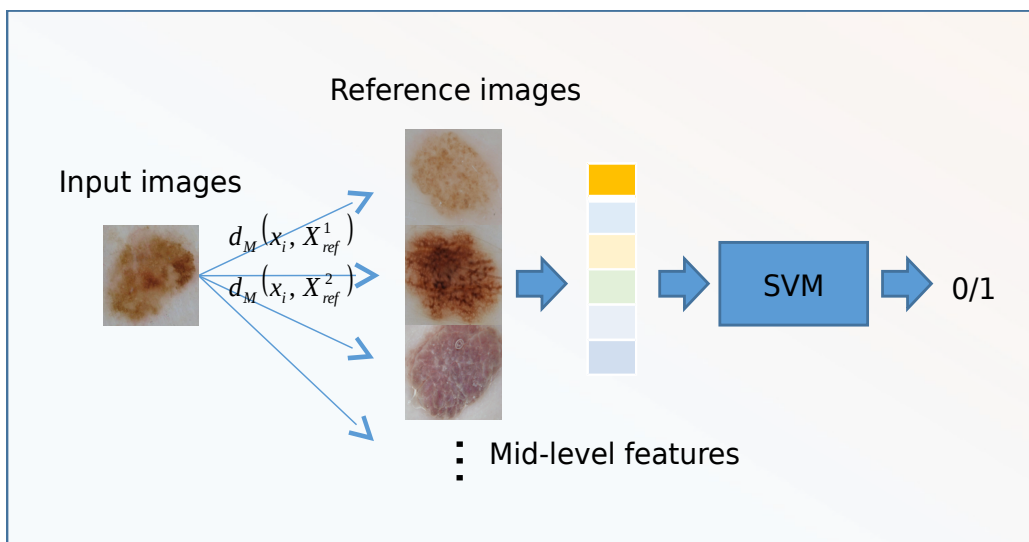
$$P(M) = \sum_{x_i, x_j \in V} w_{ij} f_M(x_i, x_j), \quad (4.4)$$

where w_{ij} is a weighting factor for each training pair. Instead of using a fixed weight for each pair of inputs, we update w_{ij} according to its difficulty of training the input pair (x_i, x_j) . Especially, we only focus on the violating pairs and give higher weights to those who violate the rules more. A violating pair is defined as a pair of samples that violates the learning rule. For instance, if the distance between two samples of the same class is larger than μ , the two samples are regarded as a violating pair. Similarly, two samples are also regarded as a violating pair if their distance is smaller than μ and they are from two different classes. Examples of violating pairs are shown in Figure 4.5. In this paper, we use V to denote the collection of violating pairs. The value of w_{ij} is initialized as 1 and it is updated in each training iteration according to:

$$w_{ij}^\tau = \mathcal{N}(|d_{M_{\tau-1}}(x_i, x_j) - \mu|), \quad (x_i, x_j) \in V \quad (4.5)$$



(a) Schematic plot of the distance metric learning algorithm



(b) Schematic plot for the skin lesion classification method

Figure 4.5: Schematic plot of the proposed method.

in which τ is the number of iterations. $\mathcal{N}(\cdot)$ is the normalization process, which is min-max normalization in this paper.

The learned distance values among different samples vary dramatically, some violating pairs' distance values are an order of magnitude different compared with the others. These violating input pairs are known as the hard samples (difficult cases), which are mainly caused by the appearance variation within the same class. Directly normalizing the distance values using standard min-max normalization will make the majority of weights w_{ij} be closer to 0, which indicates that the algorithm will only use these hard samples. Therefore, min-max normalization at a cutoff distance value is performed to normalize the distance difference values $|d_{M_{\tau-1}}(x_i, x_j) - \mu|$ to a fixed range $[0, 1]$. In this paper, the cutoff value is decided automatically by calculating the cumulative histogram of the distance differences. For the cumulative histogram, the y-axis of a bin represents the percentage of observations that are smaller than a specific value (x-axis of the bin). We use the bin value that accounts for 97% as the cut-off value, and distance values that are larger than the cutoff value is set to be 1. In this case, the weights are updated dynamically in each iteration and hard violating samples are given more importance during training.

The objective function in Eq. 4.4 can be solved using APG (*Accelerated Proximal Gradient*) algorithm [39], [42] by using Taylor series and approximating it quadratically with Lipschitz coefficient α at $M = M_\tau$:

$$\tilde{P}(M, M_\tau) = P(M_\tau) + \langle \nabla P(M_\tau), M - M_\tau \rangle + \frac{\alpha}{2} \|M - M_\tau\|_F^2, \quad (4.6)$$

in which $\tilde{P}(M, M_\tau)$ is the quadratic approximation of $P(M)$ at $M = M_\tau$. Expression of Eq. (4.6) can be rewritten as follows:

$$\tilde{P}(M, M_\tau) = \frac{\alpha}{2} \|M - Q_\tau\|_F^2 + P(M_\tau) - \frac{1}{2\alpha} \|\nabla P(M_\tau)\|_F^2, \quad (4.7)$$

where $\nabla P(M_\tau)$ is the gradient of $P(M)$, and $Q_\tau = M_\tau - \alpha^{-1} \nabla P(M_\tau)$. Assume M_τ is the value of M at iteration step τ , then the optimization problem is seeking the answer to the following problem at each iteration τ :

$$\min \tilde{P}_\tau(M, M_\tau), s.t. M \geq 0 \quad (4.8)$$

Steps for solving the problem using APG algorithm are provided in Algorithm 1.

Algorithm 1 APG Algorithm for Eq. 4.4

Require: X : original data;

Ensure: M

- 1: initial $M_\tau = I$, $\tau = 0$, $\alpha = \alpha_0$ and $t_0 = t_{-1} = 1$;
 - 2: **repeat**
 - 3: $Y_\tau = M_\tau + \frac{t_{\tau-1}-1}{t_\tau}(M_\tau - M_{\tau-1})$
 - 4: **repeat**
 - 5: set $Q_\tau = Y_\tau - \alpha^{-1} \nabla P(M_\tau)$; $M_\tau = (M_\tau + M_\tau^T)/2$;
 - 6: update $M_{\tau+1} \leftarrow U \text{diag}(\sigma - \frac{1}{\alpha})_+ U^T$; where
 - 7: $U \text{diag}(\sigma)$ is the EVD of Q_τ ;
 - 8: set $\alpha = \eta \alpha$;
 - 9: **until** $P(M_{\tau+1})_+ \leq \tilde{P}(M_{\tau+1}, M_\tau)$;
 - 10: $\tau = \tau + 1$; $t_{\tau+1} = \frac{1 + \sqrt{1 + 4t_\tau^2}}{2}$;
 - 11: **until** Maximum iteration number or convergence achieves.
-

After learning the metric M , a new feature representation is obtained by using the similarity information among a reference set. In this paper, the validation set is used as the reference set. Let $X_r \in \mathbb{R}^{p \times N_r}$ denote feature representations of the reference set. N_r is the number of images in the reference set, and X_r^j is the j th column of X_r , which represents the feature vector of j th image in the reference set. For a feature vector x_i coming from training or testing set, its corresponding new feature representation v_i is obtained by calculating its distance with all samples in the reference set.

$$v_i = \{d_M(x_i, X_r^1), d_M(x_i, X_r^2), \dots, d_M(x_i, X_r^{N_r}),\} \quad (4.9)$$

After v_i is calculated, "L2" normalization is performed. The dimension of the new feature space equals to the number of samples in the referent set N_r , which is 150 in this paper.

4.1.4 Classification Using SVM

In this work, SVM with the Radial Basis Function (RBF) kernel is used for classification, which is a common choice due to its good generalization ability and competing performance [59]. There are two parameters that need to be

tuned for the RBF kernel SVM, the parameter C , which is known as the capacity constant, the parameter g for the RBF kernel, which is a multiplier for the squared Euclidean distance between the two feature vectors. Details about the parameter selection can be found in Sec. 4.3.3.

4.2 Dataset

For performance evaluation, we have used the dataset from ISIC 2017 for skin lesion detection [16], which is a very challenging dataset for skin lesion classification. There are 2000 images in the training set, including 374 melanoma, 254 seborrheic keratosis, and 1372 benign nevi. The validation dataset contains 150 images and the final testing dataset contains 600 images. All the images are of various resolutions, ranging from 767×1022 to 4499×6748 pixels. Severe illumination variation, noise and various artifacts are also witnessed in this dataset.

4.3 Experimental Results

4.3.1 Platform Information

The skin lesion segmentation method is implemented using Keras while the classification method is implemented using Matlab. All the experiments were conducted on a desktop with Intel(R) i7-7700 4.2 GHz CPU and a GPU of Nvidia GeForce GTX 1080Ti with 11 GB memory.

4.3.2 Parameter Selection

There are two parameters of SVM that need to be tuned for the proposed skin lesion classification method. The best parameters C and g of the proposed method are selected by conducting 5-fold cross-validation on the training dataset using the deep features extracted by ResNet. Note that 2 binary classifiers for the classification Seborrheic Keratosis (SK) and Melanoma (MEL), are trained in the skin lesion classification task, therefore, we select different sets of parameters for different tasks. The best parameters used in this paper

are shown in Table 4.1. In addition, the best parameters for the proposed method without MFL are also provided.

Table 4.1: Parameters used for performance evaluation.

	Proposed		without MFL	
	C	g	C	g
SK	1	0.001	0.5	0.0156
MEL	2	0.002	0.5	0.0313

4.3.3 Comparison with Features Extracted via Pretrained CNN

To show the advantages of the learned mid-level feature representation, we first compare the learned mid-level features with the raw features obtained by the pretrained CNN models. We achieve this goal by comparing the proposed method with and without the MFL on the same test data. The best parameters of the proposed method without the MFL module have also been selected to make a fair comparison. Test data augmentation is used to increase performance as in previous work [24], [50]. In addition, to show the influence of the input features, we provide experimental results with input features extracted via the pretrained ResNet and DenseNet, respectively. Dimensions of the extracted features for ResNet and DenseNet are 2048 and 1920, respectively. After applying PCA (with 99% energy preserved), the reduced dimensions are 700 and 532 correspondingly. The results are shown in Table. 4.2 and Table 4.3.

Table 4.2: Comparison of the proposed method with and without MFL module.

Networks		ResNet			
Metric		AUC (in %)		ACC (in %)	
Task	Method	without	with	without	with
		MFL	MFL	MFL	MFL
Mel		77.67	84.29	81.17	84.33
SK		91.00	93.71	87.00	90.17
Average		84.34	89.00	84.09	87.25

Table 4.3: Comparison of the proposed method with and without MFL module.

Networks		DenseNet			
Metric		AUC (in %)		ACC (in %)	
Task	Method	without MFL	with MFL	without MFL	with MFL
	Mel		76.34	84.29	81.17
SK		93.32	93.42	87.83	89.67
Average		84.83	88.85	84.50	87.00

As observed in Table 4.2 and Table 4.3 , the learned mid-level features consistently outperform the raw features extracted by pretrained CNN. Features extracted via ResNet achieve comparable performance with the features extracted via DenseNet. Especially, for the ResNet features, the proposed method with MFL module achieves 4.7% higher for the average AUC score and 3.2% higher for the average ACC score compared with the proposed method without MFL module. For the features extracted by DenseNet, the proposed method with mid-level features as input outperforms the one with original features as input by 4.0% and 2.5% for the average AUC and ACC scores. Experimental results show that the mid-level features can significantly improve the performance, this is because additional discriminative power is gained using mid-level feature learning.

To visualize the distribution of raw features obtained using pretrained ResNet and corresponding mid-level feature representation, t-SNE is used to visualize the high-dimensional data following [50]. The t-SNE first reduces the dimension of original features to 50 by PCA (for speed up), and then to 2 by using the t-SNE Barnes-Hut algorithm [49]. It allows us to visualize the cluster of high-dimensional data to some degree. The Visualization plot is shown in Figure 4.6. From Figure 4.6, we can see that the raw features are more likely to mix together (for both the training and testing data), especially for the samples of melanoma and nevus. This means the raw features have limitations in dealing with those hard samples. The mid-level features learned from pretrained ResNet show obvious grouping behavior for the training data. Samples are more likely to cluster together if they are of the same class, and

hence the three-class skin lesions become more discriminative after using MFL. For the mid-level features of testing data, the melanoma and nevus become more distinguishable, although not completely separable. Some hard samples are identified during the learning phase.

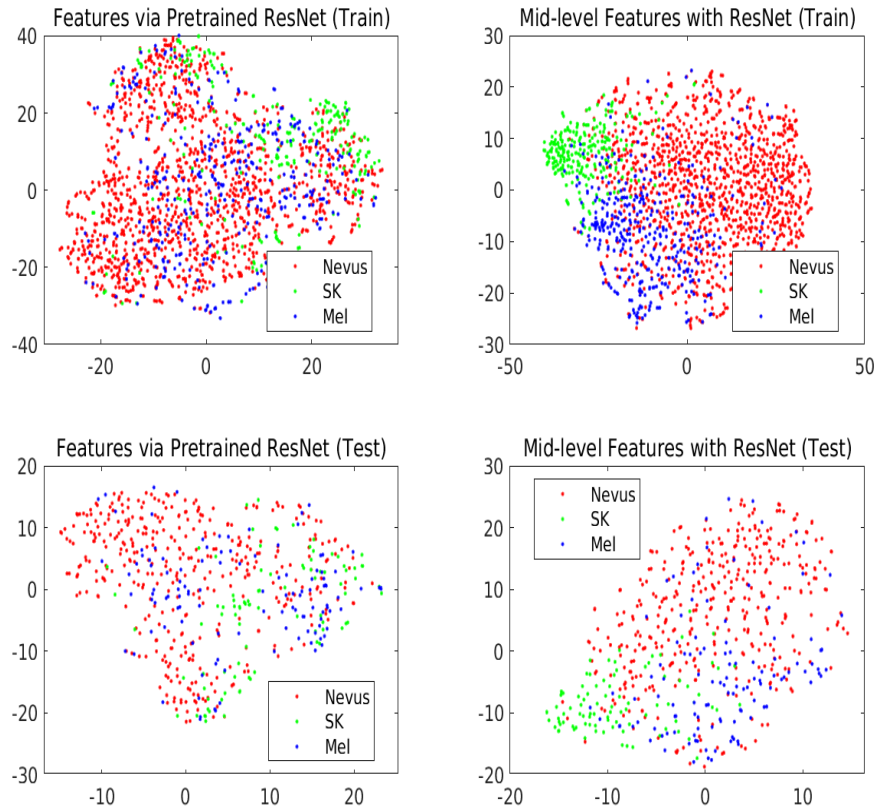


Figure 4.6: t-SNE visualization of the raw features extracted via pretrained ResNet and the learned mid-level features given raw features obtained from pretrained ResNet. The first and second row show scatter plots of the training data and testing data, respectively.

4.3.4 Comparison with Features Extracted via Finetuned CNN

In this section, we first finetune the pretrained ResNet and DenseNet for the classification tasks by changing the output dimensions of the last fully connected layers to be 2 (i.e., the number of classes). For each neural network, we set the batch size to be 24 and train it for 100 epochs. Adam optimization

Table 4.4: Comparison of the proposed method with the finetuned CNN and finetuned features using AUC scores (in %).

Method	MEL	SK	Avg
ResNet	77.30	94.19	85.75
ResNet + SVM	80.92	93.94	87.43
Proposed (ResNet)	84.29	93.71	89.00
DenseNet	84.52	92.51	88.52
DenseNet + SVM	84.66	90.62	87.64
Proposed (DenseNet)	84.29	93.42	88.85

algorithm with a learning rate of 0.0001 is used. The best models are selected by the validation performance. Data augmentation techniques including random resize cropping (70% to 100% of the original size), random horizontal and vertical flipping, random rotation (-20° to 20°) and normalization are used. Afterward, finetuned features are extracted and SVM classifiers are trained in order to compare the mid-level features with the finetuned CNN features. The same steps described in Sec. 4.3.3 are used but with finetuned CNN features as inputs. It is worth noting that parameter selection and test augmentation are also performed to make a fair comparison.

Experimental results are shown in Table 4.4. Method "ResNet" is the finetuned ResNet for classification. "ResNet + SVM" method uses the SVM to classify the features extracted via finetuned ResNet. "Proposed (ResNet)" is the proposed method, which uses features extracted via pretrained ResNet. The same definition is used for DenseNet. As shown in Table 4.4, the proposed method achieves the best average performance for both features extracted by pretrained ResNet and DenseNet. The proposed method (ResNet) outperforms the finetuned ResNet by 3.25%, and ResNet + SVM by 1.57%. The three methods obtain similar performance when using the DenseNet models. The proposed method outperforms the finetuned DenseNet by 0.33%, and DenseNet + SVM by 1.21%. The best performance for seborrheic keratosis classification is obtained when using the finetuned ResNet (94.19%). For the melanoma classification, the best performance is obtained by using SVM over finetuned DenseNet (84.66%).

4.3.5 Comparison with State-of-the-art Methods

In this section, the proposed approach is compared with the state-of-the-art methods. Due to the fact that ISIC 2017 dataset is a challenge dataset, tricks such as ensemble are widely used among the existing methods. Prevalent methods get the final performance by fusing outputs from different trained neural networks. Here we give a brief introduction about the compared methods in Table 4.7: [53] trained ResNet-50 with different optimization methods, and selected the best combination of fine-tuned CNN through cross validation. Besides, a manual decision rule with metadata (age, sex information) is also adopted. [54] used ResNet-101 and Inception-v4 models. The final results were obtained by ensembling 7 trained neural networks with a meta learning model to assemble these models. [10] fused outputs of the binary ResNet and 3-class ResNet to get the final results. [50] used AlexNet, VGG16, ResNet-18 and ResNet-101 models. Extensive models are used to boost performance. The final results of a single architecture (e.g., ResNet-18) were acquired from 18 different models (obtained by different training settings). [91] used multi-task framework (GoogleNet and U-net) for learning skin lesion segmentation and classification jointly. [24] trained a FCN for detecting ROI. In addition, [24] also incorporates the meta-data information and attribute information to improve performance.

In the proposed method, ensemble method is used during the testing time to improve the performance. Our final model is obtained by fusing outputs given input images from multiple scales (based on performance on the validation set), which does not require extra training process. Performances of inputs with different scales are shown in Table 4.5. As shown in Table 4.5, for input images with different scales, the best performance is obtained at scale 224. Scale 672 comes the second best, and scale 448 gives the least satisfactory performance.

Table 4.5: Average AUC of the proposed method with different input sizes.

Scale	ResNet	DenseNet	Fusion
$I_{[224,224]}$	89.00	88.85	90.67
$I_{[448,448]}$	88.22	88.13	89.35
$I_{[672,672]}$	89.04	88.28	89.40

Experimental results regarding the ensemble of scales on the validation set are shown in Table 4.6. From Table 4.6, we can see that, in general, adding more scale information can improve the performance, but the best performance (AUC of 95.2%) is obtained by fusing results with input scale 224, *224_entire* (i.e., the whole image without ROI segmentation), and 672. The fusion with scale 448 gives less satisfactory performance compared with the other scales. This is consistent with the results reported in Table 4.5 that scale 448 performs the least satisfactory and fusing a less satisfactory output (obtained with input scale 224) would not increase the performance. This may be due to the fact that the parameters of SVM are not selected for these inputs with different scales. When the input size is larger than $[224 \times 224]$, the output of pretrained neural networks will be multi-channel features instead of one feature vector. We reshape the multi-channel features into one feature vector, and PCA is then used to reduce the feature dimension, which is described in Sec. 4.1.2. This will result in raw features with different input dimensions, given the inputs of different scales. The finetuned parameters of scale 224 are used for the proposed method for simplicity. The use of entire images is to get information about the lesion size and skin regions, which can benefit the proposed model [12]. Therefore, our final performance is obtained by fusing the outputs with input scale $[224 \times 224]$, $[672 \times 672]$ and the entire image with scale $[224 \times 224]$. It is worth noting that a consistent ensemble trend regarding the fusing of different input scales has also been found on the testing set.

Table 4.6: Ensemble performance with input of different scales on the validation set.

224	224_entire	448	672	AUC
✓	✓			94.0
✓		✓		93.6
✓			✓	94.1
✓	✓	✓		94.5
✓	✓		✓	95.2
✓		✓	✓	94.3
✓	✓	✓	✓	94.9

Final results compared with state-of-the-art methods are shown in Table 4.7. The column "ensemble" indicates whether the compared methods use ensemble technique or not, and the column "external data" shows the number of external data used for training neural networks. The external data plays a vital role in the training of CNN models for the skin lesion classification task. For instance, [24] got the best performance of 90.8% (vs. 91.7%) for the models with fewer external training sets, even when it incorporated the meta-data and attribute data. From Table 4.7, we can see that the proposed method provides superior performance compared with state-of-the-art methods without using external data. The proposed method achieves the best AUC of 97.1% for seborrheic keratosis, which verifies the effectiveness of the proposed method.

Table 4.7: Performance comparison with state-of-the-art methods on ISIC 2017 dataset (AUC score).

Method	ensemble	external data	MEL	SK	Avg
[53]	Y	1444	86.8	95.3	91.1
[54]	Y	7544	87.4	94.3	90.8
[10]	Y	1600	87.0	92.1	89.6
[91]	N	0	83.0	94.2	88.6
[24]	N	2828	87.3	96.2	91.7
[50]	Y	187	87.3	95.5	91.4
Proposed	Y	0	87.0	97.1	92.1

We also display some challenging images that have been correctly classified by the proposed method in Figure 4.7. The left images are skin lesions of melanoma, while the right images are skin lesions of seborrheic keratosis.

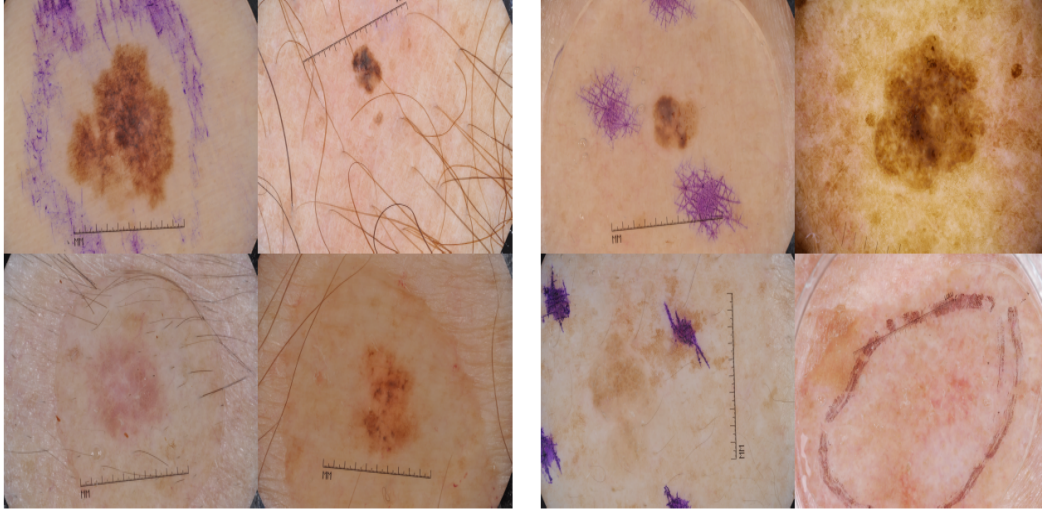


Figure 4.7: Examples of correctly classified images: left: melanoma; right: seborrheic keratosis.

Strong visual similarities and artifacts are observed in these two types of images, yet the proposed method successfully classifies these hard samples, which implies that the proposed method can tackle difficult samples.

4.4 Discussions

4.4.1 Segmentation Performance

In this section, we compare the proposed segmentation model with the U-net [67] and SegNet [7]. In order to make a fair comparison, the inputs of these neural networks are the same. Experimental results can be found in Table 4.8. Proposed_1 and Proposed_2 represent the proposed method without a test ensemble and with a test ensemble technique, respectively. As we can see from Table 4.8, the proposed method obtains the best performance, which implies that the adoption of dilated convolution layers can significantly improve the performance. Meanwhile, Proposed_2 outperforms Proposed_1, which shows the importance of test data ensemble.

Table 4.8: Comparison of different methods on ISIC 2017 skin lesion segmentation dataset.

Method	JA	DC	ACC	SEN	SPE
SegNet [7]	0.700	0.797	0.917	0.815	0.974
U-net [67]	0.686	0.788	0.915	0.755	0.969
Proposed_1	0.740	0.830	0.926	0.828	0.965
Proposed_2	0.752	0.840	0.930	0.829	0.988

4.4.2 Effect of Lesion Segmentation

To determine the influence of the skin lesion segmentation, we conduct experiments with and without skin lesion segmentation as a primary step, and experiment results are shown in Table 4.9.

I_{whole} means using the whole images as inputs, and I_{ROI} means using the ROI images obtained with skin lesion segmentation method as inputs. As shown in Table 4.9, the proposed method obtains significant improvement if we use the ROI images as input. This is mainly because the interest regions of skin lesions are of various scales, and some targets are very small and only occupy a small region of the entire skin lesion. Directly down-sampling all images to the same size ([224, 224]) will lose detail information about skin lesions, and make it even difficult to observe patterns of skin lesions. Some examples of the predicted binary masks are shown in Figure 4.8.

Table 4.9: Average AUC of the proposed method with and without skin lesion segmentation.

Input	ResNet	DenseNet	Fusion
I_{whole}	84.11	84.62	86.10
I_{ROI}	89.00	88.85	90.67

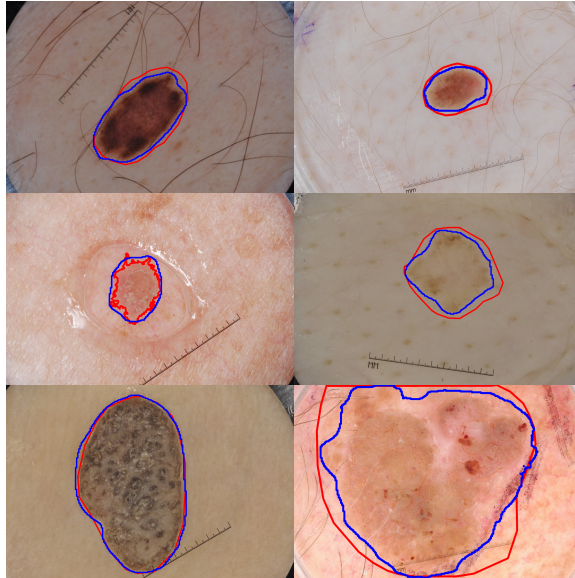


Figure 4.8: Segmentation results of the proposed method. The red contours are the ground truths, and the blue contours are the segmentation results.

4.4.3 Effect of Weighting Factor for APG Algorithm

In this section, experiments with different weighting schemes are conducted to show the benefits of the proposed approach. Specifically, we compare the proposed method with uniform weights, and the experimental results are shown in Table 4.10. The weighting scheme used in this paper can get an improvement of 1.12% and 0.15% with features extracted via ResNet and DenseNet, respectively. Though minor improvement is observed for the features extracted via DenseNet with online weighting, it can still be regarded as useful overall. One possible reason is that focusing more on hard violating pairs can benefit the training phase of metric learning.

Table 4.10: Effect of weighting factor for APG algorithm.

Weighting	ResNet	DenseNet
Uniform	87.88	88.60
Proposed	89.00	88.85

4.4.4 Advantage of Soft Discriminative Feature

As shown in Eq. 4.2, the metric learning problem can also be regarded as a discriminative subspace learning problem. The new feature representation can be represented as Lx_i given the input feature x_i , which is a discriminative feature representation. Compared with the discriminative feature Lx_i learned based on metric learning, the proposed mid-level feature representation v_i in Eq. 4.9 is a soft descriptor, which uses affinity information as the new feature representation. In this section, we also implement experiments to compare the proposed mid-level features v_i with the discriminative feature Lx_i . To make a fair comparison, the best parameters of the discriminative features based on metric learning are also selected as described in Sec. 4.3.3. Experimental results are shown in Table 4.11.

Table 4.11: Comparison of AUC scores of the proposed method using discriminative features and mid-level features.

Input	ResNet	DenseNet	Fusion
Discrim Fea	86.53	87.68	88.71
Mid-level Fea	89.00	88.85	90.67

From Table 4.11 we can see that the mid-level features outperform the discriminative features for both the features extracted via ResNet and DenseNet. This is because learning an optimal feature representation that can well separate all the samples (especially the hard samples) is very difficult. In contrast, the proposed mid-level feature representation is a soft discriminative descriptor, where the relationships of visual similarities and distinctions can be kept for some difficult cases (hard samples) as long as the remaining relationships are captured correctly. Also note that, compare with Table 4.2 and Table 4.3, the discriminative features obtain better performance than the original features, which demonstrates that the original features have poor discriminative power, and using the discriminative features can promote the classification task.

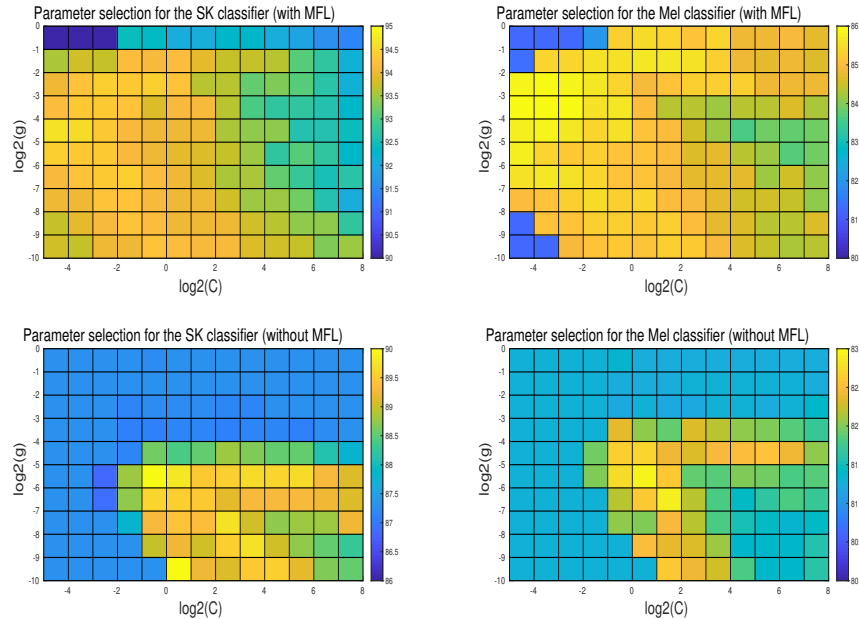


Figure 4.9: Performance of the proposed method (with and without the MFL module) on the parameter space. The learned mid-level features are more robust and discriminative compared with the original features.

4.4.5 Robustness of the Proposed Mid-level Features Against Parameters

In this section, to show the robustness of the proposed mid-level feature representation against parameters, performance (5-fold cross-validation on the training set) of the proposed method on the parameter space is given in Figure 4.9. As shown in Figure 4.9, the mid-level features outperform the original features by a large margin for both binary classifiers. Especially, the AUC of proposed method for the SK classifier ranges from 90% to 95% across the parameter space, while the AUC of proposed method without MFL ranges from 86% to 90%. The best performance of the proposed method is about 5% higher than the proposed method without MFL, which proved that the mid-level features contain more discriminative power. A similar trend is also observed for the MEL classification. Also note that the ranges of color bars are similar for the method with and without MFL, and as the score increases, the color of the parameter space changes from blue to yellow. A large area of the

proposed method’s parameter space is yellow while only a minor part is yellow for the proposed method without MFL, which demonstrates the robustness of the proposed mid-level features.

4.4.6 Time Complexity

There are two binary classification tasks in this study. For simplicity, the following time complexity is calculated on the classification task of seborrheic keratosis using features extracted via ResNet. The platform information has been described in Sec. 4.3.1. The training time of the proposed method is a total of 96 minutes. Out of 96 minutes, 92.4 minutes are spent to train the segmentation network, and about 3.6 minutes are used to train the classification model. The testing time is 0.39 seconds on average for one given image. Typically, the extra time induced by the proposed method (mid-level feature processing) is 0.13 seconds, which is relatively fast. The fast inference time of the proposed method indicates its potential in clinical application.

4.5 Conclusions

Automatic melanoma detection is a challenging task due to the large inter-class similarity and intra-class variation, and complex skin conditions among different skin lesions. In this paper, a novel framework for skin lesion classification is proposed. Skin lesion segmentation is first performed to get the ROI images for the later classification task. A novel mid-level feature representation is obtained by using metric learning and a reference set. The learned mid-level feature representation contains affinity information among image samples, which is a soft discriminative feature, having more tolerance to the hard samples and thus being more robust. Experimental results show that skin lesion segmentation can benefit the subsequent classification task. Meanwhile, the learned mid-level features obtain much better performance compared with the original features. Experimental results have verified that the proposed approach outperforms state-of-the-art CNN based methods, which verify the effectiveness of the proposed method.

Chapter 5

Integration of Light Scattering with Machine Learning for Label Free Cell Classification

In the previous chapters, high-resolution RGB images (dermoscopic images) have been used to detect skin cancer. In this chapter, a different type of medical modality (single-cell angular light scattering patterns) is used to provide a way for cell identification and analysis. Typically, it can be used for the identification of any type of cell once useful signatures have been obtained based on the scattering patterns. Previous studies have shown that machine learning techniques based on scattering patterns can successfully detect the malignant cells from healthy cells [41], [78], [81]. In this chapter, a more challenging task, the identification of cells of different states, is introduced.

Conventional bench-top flow cytometer, also known as the FACS, is a valuable tool for cell identification in many biological and health-related applications [62]. It is based on the measurement of fluorescence of molecules that are attached to the illuminated target cells [62]. However, these fluorescent biomarkers can interfere with the function of the cells they bind to, hindering further potential analyses and complicating interpretation [30]. Additionally, adequate biomarkers are not available for many cell types, and they can be

An extended version of this chapter has been published as W. Yu*, L. Liu*, X. Liu, W. Wang, M. Z. Islam, C. Dong, C. R. Garen, C. R. Woodside, M. T. Gupta, M. Mandal, W. Rozmus, Y. Y. Tsui, "Integration of light scattering with machine learning for label free cell detection," *Biomedical Optics Express*, vol. 12, no. 6, pp. 3512-3529, 2021. (* co-first author)

difficult or expensive to develop for organisms such as microbes and viruses [30]. The development of label-free techniques for cell identification is thus of high interest to scientists to avoid such complications.

The cell light scattering patterns are dominated by the small-scale 2D speckle patterns originating from light scattering from the mitochondria as they have a higher refractive index than other organelles [71], [76], [78], [79]. These patterns are different due to the variation in the shape, number and distribution of the mitochondria. Light scattering has been studied as a label-free technique for single-cell analysis by several research groups [60], [68], [87], [95]. In earlier studies in our group, we have used angular light scattering patterns as signatures for label-free cell identification [40], [71], [76], [77], [79]. This technique was applied to study yeast, human Raji [76] and hematopoietic stem cells [71]. Single cells were identified by comparing the experimental 2D scattered light patterns measured using a Charged-coupled Device (CCD) camera with the numerical simulation results. In these studies, two statistical features, including the number and average area of speckles, were used to distinguish the cells. The simulations were carried out by solving the Finite-Difference Time-Domain (FDTD) technique with simplified models for optical properties of the cells [40]. The cells have been defined in simulations as three-dimensional dielectrics of spherical or oval shapes, containing different cell organelles of varying indices of refraction [71], [78]. Numerical simulations identify the main scattering centers in cells such as nuclei, small organelles, e.g., mitochondria or lysosomes, and reproduce the characteristic features of the experimental angular scattering spectra.

Statistical study of the speckles' spatial distribution in the scattered light patterns allows us to effectively distinguish one cell type from the other. In this chapter, laser scatter patterns of two groups of SH-SY5Y neuroblastoma cells, including the ST and NT cells, are collected. The SEM images for the ST and NT SH-SY5Y cells were taken to better understand their morphology, which are shown in Fig. 5.1. As shown in Fig. 5.1, there is an obvious difference between SEM images of the ST and NT SH-SY5Y cells. The NT cells in Fig. 5.1 are characterized by a relatively smooth surface and quasi-spherical shapes,

while the ST cells have rough and almost sponge-like surfaces with various shapes. Despite such different physical features that are displayed in Fig. 5.1, the apparent similarities between the scattering patterns of these two kinds of cells add difficulties in identifying the main components contributing to the scattered light spectra. The simple analysis for cell identification proposed in the previous work [71], [78] is proved insufficient in discriminating these two groups of cells, and more sophisticated techniques are needed. In this study, DL methods have been developed for the evaluation of 2D scattered laser light patterns for cell identification. More details about the proposed method are introduced in the following sections. The objective is to classify the patterns into two classes: ST and NT cells.

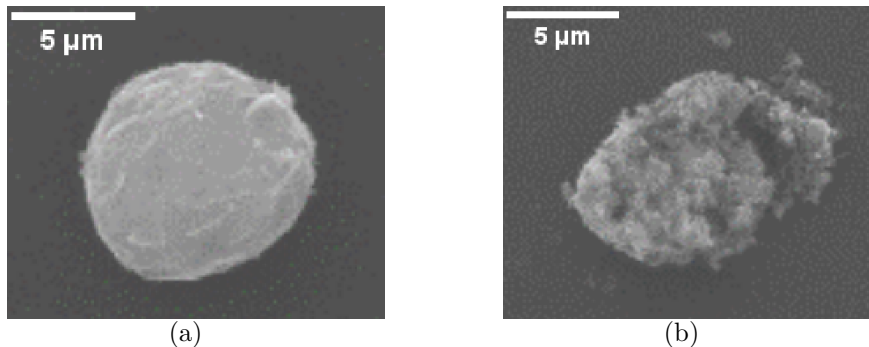


Figure 5.1: SEM images of non-treated (left) and staurosporine-treated (right) SH-SY5Y cells. The two images correspond to cells of the similar size approximately 8 μm in diameter.

5.1 Experiment Setup

Because it is relatively straightforward to obtain side scattering patterns experimentally, most of the previous experimental studies [41], [45], [76], [77], [80] collected scattered light from single cells in the side direction with a single probing wavelength. A label-free cell identification technique pioneered by University of Alberta researchers has been developed for the identification and analysis of any cells, which can collect scattering patterns in three directions (forward, side, and backward) simultaneously. Both red laser and green laser have been used to analyze the effect of wavelength. It is compatible with au-

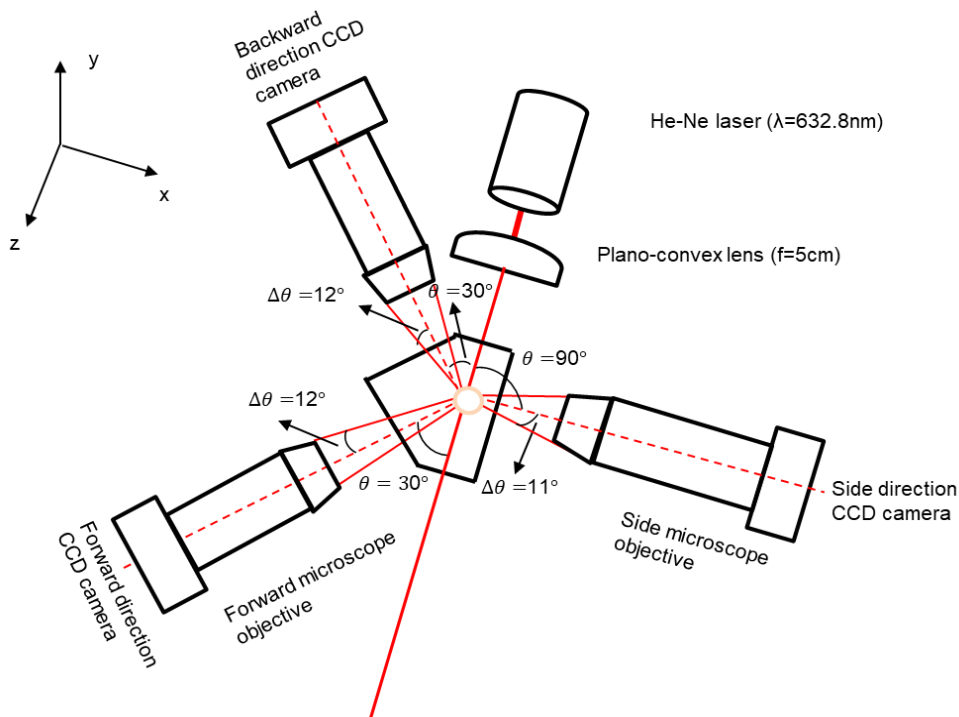


Figure 5.2: A schematic diagram of the experimental setup used to obtain two-dimensional light scattered patterns of cells in three directions.

tomation technology, and if it is successfully developed, it will overcome many of the drawbacks of the conventional FACS flow cytometry technology. The technique identifies single cells by identifying their 2D scattered laser light patterns measured using CCD cameras. A schematic diagram of a typical experimental setup for collecting 2D scattered laser light patterns from cells is shown in Fig. 5.2 [71]. Key components of the system include a probing laser, a sample holder, and CCD cameras with microscope objectives. Each microscope objective and CCD camera were connected by a tube, and the system was placed on three-directional translation stages. The central lines of the forward, side, and backward microscope objectives were 30° , 90° , and 150° away from the laser beam direction, respectively. The ranges of light collection angles were between 18° and 42° in the forward direction, between 79° and 101° in the side direction, and between 141° and 159° in the backward direction, with respect to the laser beam direction (z -direction). The overlap of the laser beam and the observation region of each CCD camera with micro-

scope objectives defines a small measurement volume of approximately 0.002 mm^3 . Normally a cuvette that contains a dilute cell solution is used as the sample. The cell solution is diluted to a concentration typically around 2000 cells per ml to ensure single-cell measurement. The laser scattered light from a single cell is captured by the microscope objectives and CCD cameras for further analysis.

Scattering patterns from the red laser illumination are shown in Fig. 5.3 and Fig. 5.4, while those from green laser illumination are shown in 5.5 and Fig. 5.6. The top, middle, and bottom rows show scattering patterns measured in forward, side, and backward directions, respectively. As we can see from these scattering patterns, it is very difficult to discriminate between the ST and NT cells by human eyes.

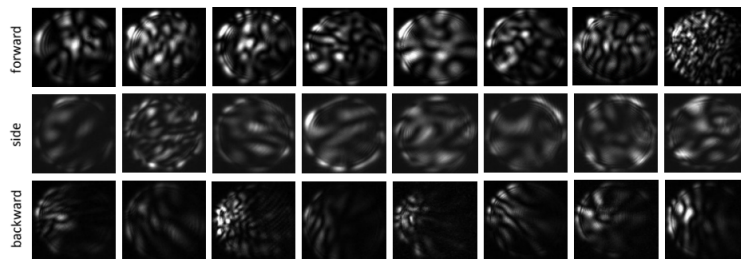


Figure 5.3: Scattered light patterns from treated SH-SY5Y with red laser. The angular ranges corresponding to the three directions of measurements: 18° - 42° (forward), 79° - 101° (side), and 141° - 159° (backward).

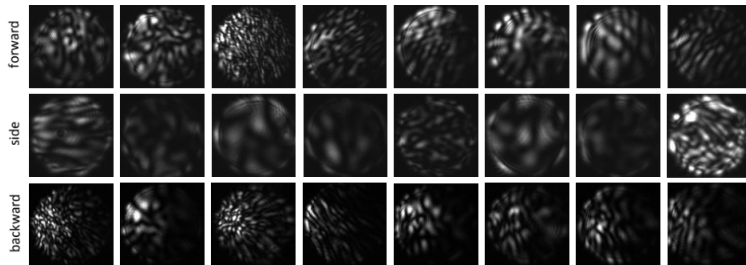


Figure 5.4: Scattered light patterns from non-treated SH-SY5Y with red laser. The angular ranges corresponding to the three directions of measurements: 18° - 42° (forward), 79° - 101° (side), and 141° - 159° (backward).

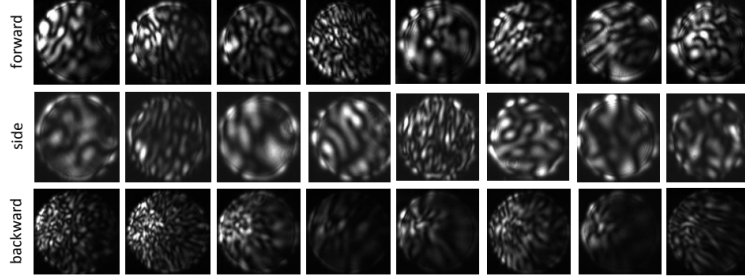


Figure 5.5: Scattered light patterns from treated SH-SY5Y with green laser. The angular ranges corresponding to the three directions of measurements: 18° - 42° (forward), 79° - 101° (side), and 141° - 159° (backward).

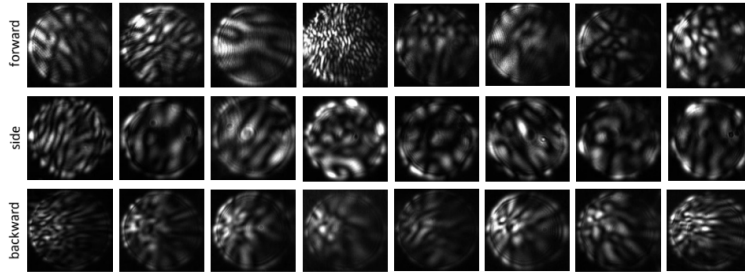


Figure 5.6: Scattered light patterns from non-treated SH-SY5Y with green laser. The angular ranges corresponding to the three directions of measurements: 18° - 42° (forward), 79° - 101° (side), and 141° - 159° (backward).

5.2 Proposed Method

In this section, we present the machine learning technique used in this chapter for the classification of scattered light patterns. Since the dataset is small, transfer learning based on CNN is used. In transfer learning, a model trained on a large dataset is used, and its knowledge is transferred to a smaller dataset. It has been widely used as a feature extraction method. The early convolutional layers of the pretrained CNN are frozen, and the features at a certain layer are used as the feature representation.

In the experiment, we use pretrained DenseNet as the offline feature extractor [31]. Typically, DenseNet-201 [31] is used which consists of 201 layers, including convolutional (Conv), ReLU activation (ReLU), batch normalization (BN), pooling layers, etc. It is named as DenseNet since dense connections between all layers are utilized, which are achieved by merging all the previous layers as the inputs of the later layers. The overall schematic of the ML mod-

ule is shown in Fig. 5.7. The arcs indicate the dense connections between two layers.

The input images are resized to 224x224 pixels and are fed to a pretrained CNN for feature extraction. A feature vector of length 1920 is obtained at the GAP layer of the DenseNet for each image. The feature vectors are then fed to an RBF-kernel SVM that outputs a real value score in the range $[0, 1]$. The score determines the probability of the image class (i.e., the probability of being an NT or ST). As the pretrained DenseNet is used only as a feature extractor, there is no training required for the DenseNet, and only the SVM needs to be trained to obtain the classification output. There are two parameters c and g that need to be optimized for the kernel SVM. In this study, the best parameters are automatically selected by 5-fold cross-validation using the training data.

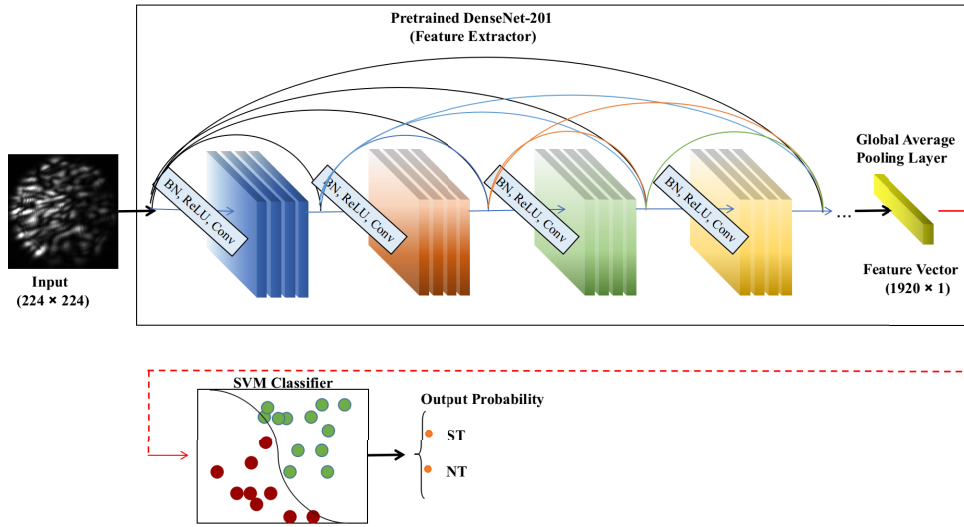


Figure 5.7: Schematic of the machine learning module.

5.3 Dataset

Experimental scattered light patterns of both the ST and NT SH-SY5Y cells in all three directions have been successfully obtained using our experimental setup in Fig. 5.2. In this section, we did not use the scattering images from the backward direction, as it is difficult to collect the scattering images, which

are usually of low-intensity values. Scattering patterns obtained using the red laser are used, as they can give us better performance.

To verify the effectiveness of the proposed technique for cell classification, a dataset consisting of 360 images, including 180 NT and 180 ST cell images, is used in the following experiments. All the images are resized to [224, 224] before they are fed into the CNN model. 5-fold cross-validation is performed to avoid the randomness of the train-test split. And the average performance is used as the final result.

5.4 Experimental Results

5.4.1 Platform Information

The proposed method is evaluated on a Desktop with Intel(R) i7-7700 4.2 GHz CPU with 32 GB memory and a GPU of Nvidia GeForce GTX 1080Ti with 11 GB memory. The features are extracted using pretrained DenseNet-201 in Pytorch, and the SVM classifier is implemented using Matlab.

5.4.2 Comparison with Conventional Machine Learning Methods

To analyze the performance of the proposed technique, we also compare the proposed method with other scattering analysis techniques [41], [78], [80]. Two Speckle Features (SF) including the speckle number and average area, are used for cell classification in [78]. GLCM features are used in [41], and similarly, HOG features are used in [80]. To extract SF, we first use a Gaussian filter to smooth the original images, and speckle segmentation is then performed so as to extract the number and average area of the speckles. For GLCM features, the gray level range is set to be [0,255], the pixel pair distance is set as 1, and four directions of 0° , 45° , 90° , and 135° are used to calculate the GLCM. 22 statistical features can be calculated from the GLCM in each direction. The final feature representation is obtained by concatenating the features from these 4 directions, resulting in a feature vector with a dimension equal to 88. For HOG features, we divide the original images into non-overlapping 16x16

pixel blocks, and 9 bins are used for calculating the histogram within each block. The dimension of the HOG features used in this paper is 1764 for each image. In order to make a fair comparison of these methods, all the experiments are conducted using the same experimental setup.

Experimental results are shown in Table 6.1. The hand-crafted features (Hand-Crafted) use the concatenation of the SF, GLCM and HOG features. Among all the methods, the proposed method obtains the best performance with an accuracy rate of 91.11%, sensitivity of 92.78%, specificity of 89.44% and AUC score of 96.95%, which shows the effectiveness of the proposed method. SF features perform the worst among all the features. This is because the SF is only a two-dimensional statistical feature representation, which can not provide sufficient information for characterizing different classes of cells. The GLCM and HOG features can capture the texture features of images. Over 10% of accuracy improvements are observed for the GLCM and HOG features, compared with the SF. This indicates that the texture features are more useful for scattering patterns discrimination compared with the two statistical features SF. Among the hand-crafted features, the HOG features present the best performance. The concatenation of the three features provides further improved performance since it captures more diverse and useful information. However, it still gives poorer performance compared with the proposed technique. This is because the hand-crafted features are low-level features. They need careful design given different images, and hence they are data-dependent. They perform unsatisfactorily when the task is difficult. For instance, when the visual patterns are extremely similar for different groups of cells in this case. However, the deep features can extract rich texture and semantic features. They are designed for the classification task. Therefore, they have more discriminative power. The experimental results in Table 5.2 have verified that the deep features are more powerful for scattering patterns discrimination compared with the hand-crafted features.

Table 5.1: Classification Performance of the proposed method.

Classification Method	ACC	SEN	SPE	AUC
SF [78]	61.67	63.33	60.00	64.86
GLCM [41]	71.11	70.56	71.67	77.19
HOG [80]	75.28	76.11	74.44	83.15
Hand-Crafted	77.78	82.78	72.78	86.62
Proposed ML	91.11	92.78	89.44	96.95

5.4.3 Percentage Prediction in A Mixed Solution

In some medical applications, one cares more about the percentage of a certain type of cell, which can be a diagnostic factor for some diseases. In this paper, the PTST is defined below:

$$\begin{aligned}
 PSTC &= \frac{No.ofSTcells}{TotalNo.ofcells} \times 100\% \\
 &= \frac{TN + FN}{TP + FP + TN + FN} \times 100\%
 \end{aligned} \tag{5.1}$$

To measure PTST, the same testing dataset is used but with different percentages of ST cells (selected randomly). In this way, a calibrated testing set with different percentages of ST cells is constructed. This procedure is repeated 10 times, and the average PTST value is used as the prediction result. Specifically, three experiments with ground truth PTST values of 50%, 33%, and 20%, are conducted. Experimental results for the prediction of the PTST are shown in Table 5.2. It is observed the predicted PTST value is very close to the ground truth PTST when the dataset includes the same percentage of NT and ST cells. However, when the number of NT cells exceeds the number of ST cells by a large margin, the predicted PTST value deviates slightly and is usually larger than the ground truth value. However, the predicted percentages are still within 5% error compared with the ground truth PTST values, which may be acceptable for some diseases in making diagnostic decisions.

Table 5.2: Prediction of PTST.

Testing Dataset	Ground Truth Percentage	Predicted Percentage
36 ST, 36 NT	50.0	49.1
18 ST, 36 NT	33.3	35.2
9 ST, 36 NT	20.0	24.2

5.5 Discussions

5.5.1 Comparison of the Proposed Technique Using Scattering Images From the Side Direction

The angular direction of the CCD cameras with the laser beam can affect the scattering patterns. To analyze which direction is better for the cell discrimination, we also collected a dataset in the side direction with the red laser. The same experimental setup is used to make a fair comparison, and the experimental results are shown in Table 5.3. As shown in Table 5.3, the proposed technique with laser scattering images from the side direction can obtain an accuracy of 88.89%, which is slightly worse than the scattering images collected in the forward direction. However, it is still a satisfactory performance overall. Since the induce of the staurosporine treatment can result in mitochondria change, the satisfactory experimental results also demonstrate that both the side and forward direction carry information regarding the mitochondria information. The forward direction laser scattering images are more beneficial for the cell discrimination. Hence, we use the laser scattering patterns in the forward direction as the input of the proposed technique. Combining scattering patterns from different directions of the same single cell by the fusion method can promote classification performance. However, the current experimental setup cannot guarantee which pair of scattering images (one from the forward direction and the other one from the side direction) corresponds to the same cell. The standard fusion method of the machine learning technique can not be used in this case. However, combining scatter images from different directions can provide more information, and it will be of significant meaning to analyze this in our future research.

Table 5.3: Performance comparison of the proposed technique with scattering images from the side direction and forward direction.

Direction	ACC	SEN	SPE	AUC
Side	88.89	91.67	86.11	95.98
Forward	91.11	92.7	89.44	96.95

5.5.2 Inference Time

The inference time is approximately 0.033 seconds on average for a single cell image. Feature extraction costs most of the time, which is approximately 0.0326 seconds. However, the time for feature extraction can be sped up by using some compression technology of deep neural networks or some more advanced CNN methods, which can obtain similar performance but are more lightweight and efficient. The testing time is relatively speedy, and hence the proposed technique has the potential for real-time cell classification.

5.6 Conclusions

In this chapter, a deep learning based method has been developed for the cell scattering image classification of the staurosporine-treated and non-treated SH-SY5Y cells. A label-free cytometry setup is used to obtain the scattering patterns of cells. The scattered lights are observed with CCD in the forward direction with a red laser as the incidence beam. The proposed method obtains the best performance with an accuracy rate of 91.11%, sensitivity of 92.78%, specificity of 89.44% and AUC score of 96.95%. The classification technique is then applied to predict the fraction of treated cells in a mixed solution. Experimental results show that good predict accuracy, less than 5% error compared with the ground truth, has been achieved. The inference time is very fast, about 0.033 seconds on average for a single cell image. Experimental results show that our label-free cytometry has the potential for real-time detection of different cells and may assist the fundamental apoptosis study in neuroscience research.

Chapter 6

Multi-wavelength Multi-direction Laser Light Scattering for Cell Characterization Using Machine Learning

Scales and distributions of the inner cell organelles are complex. Because of the sub-micron sizes of these inner organelles, their images cannot be well-resolved by using optical microscopy. However, information about cells, including their inner organelles, can be obtained from the light scattering patterns using the label-free cell analysis technique as described in section 1.2. In our previous study, the light scattering cell identification technique was used to distinguish blood cells based on their distinct 2D scattered light patterns in the side direction originating from their different mitochondria distributions [79]. In our recent study, as described in Chapter 5 [85], 2D scattered light patterns in the side and near forward directions probed by a red or a green laser were used to study the non-treated and staurosporine-treated SH-SY5Y neuroblastoma cells. We found that the 2D angular light scattering patterns from the near direction probing by a red laser provided the best results in distinguishing between the non-treated and staurosporine-treated cells. We speculated that the

An extended version of this chapter has been submitted to a journal as L. Liu*, M. Z. Islam*, X. Liu, M. T. Gupta, W. Rozmus, M. Mandal and Y. Y. Tsui, "Multi-wavelength multi-direction laser light Scattering for cell characterization using machine learning," 2022. (* co-first author).

main contributions of the angular light scatterings originated from their mitochondria content and cell membrane roughness. In this chapter, the machining technique is used to better understand the contributions of two factors (cell surface roughness and the number of mitochondria) to single-cell scattering patterns.

In addition to the cell properties, scattered patterns are also influenced by the observation angles of the light. Especially, scattered patterns collected in different directions reflect different cell properties. The forward scattered light that is collected in a small angular range ($\theta < 2^\circ$) is related to the refractive index and cell's size [56]. Forward light collected at larger angular range $5^\circ < \theta < 30^\circ$ contains more information regarding the cell size [86], cell nucleus and the nucleus/cell volume ratio [87]. Small internal organelles such as mitochondria contribute to the light scattering patterns at larger angles, e.g., side direction [86], [87]. The backward scattered patterns ($160^\circ < \theta < 180^\circ$) are mainly related with the cell membrane [55]. Moreover, scatter patterns collected using different wavelengths of light also carry different information about cells, which may provide complementary information and are beneficial for cell classification [81]. A systematic study about the effects of observation angles and probing wavelengths is also conducted in addition to the two cell properties. The results can also help us to better characterize and understand the morphology and mitochondria content change that may occur during cell treatment, aging, or disease development. The results can also be used to optimize our label-free cell analysis technique.

6.1 Proposed Method

Simulated data is used in this work. Lumerical FDTD software [47] was used to obtain the multi-wavelength multi-direction scattering images of the same single cell. To analyze the surface roughness and the number of mitochondria's contributions to specific scattering patterns, we have set up two sets of experiments (the simple cases and the complex cases) with parameter control. We achieve this goal by using data with one fixed parameter and one varying pa-

parameter and observing their scattering patterns and classification performance with machine learning analysis techniques. We name this experiment a simple case study. In real cases, the internal structures of cells present a small range of variations; therefore, there is certain randomness regarding the cell property. To simulate these cases, we conduct experiments by mixing scattering patterns with different parameters and observe their classification performance when the other parameter changes. For instance, we tried to classify scattering patterns into two groups with different numbers of mitochondria (75 and 100 in this study), while there are an equal number of scattering images with different roughness levels within each group. We call this a complex case. Details about the experiment implementation are introduced in the following sections. The same machine learning technique, which has been introduced in Sec. 5.2, is used in this method.

6.2 Data Generation

In the simulations, the cell models are constructed by considering different cell organelles as optical materials with scale values obtained from the previously published article [85]. The spherical shape is used for the dielectric model of the cell cytoplasm and nucleus. An elliptic shape (with minor and major axes diameters of 0.75 micron and 1 micron, respectively) is used for the mitochondria. The simulated datasets are generated with two different cell properties, surface roughness and the number of mitochondria. Since it is easy to set different numbers of mitochondria in a cell model, we only introduce the simulation of different cell surface roughnesses.

6.2.1 Numerical Simulation of Cell Surface Roughness

To introduce roughness to the cell membrane model, a computer model using a spherical dielectric with pseudo-random Gaussian fluctuations in its radius is developed. There are two parameters that control the cell surface roughness: amplitude of modulation σ and correlation length of the modulation Λ . They can control the fluctuations of the radius of a circle in a two-dimensional space.

A rough sphere is generated by first slicing the spherical cytoplasm into many circular discs in one dimension. Afterward, the roughness model is applied to each disc, and the final model is obtained by fusing these discs together. More specifically, we use h to indicate the small variation of the radius of each circular disc, such that:

$$\langle h \rangle = 0 \quad (6.1)$$

$$\langle h^2 \rangle = \sigma^2. \quad (6.2)$$

Variation (modulation) of along the periphery of a disc, h , satisfies the Gaussian distribution:

$$p(h) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(\frac{-h^2}{2\sigma^2}\right) \quad (6.3)$$

Hence, correlation function can be used to represent the occurrence of the radius modulation:

$$C(R) = \frac{h(\rho)h(\rho + R)}{\sigma^2} = \exp\left(\frac{-R^2}{\Lambda^2}\right) \quad (6.4)$$

where R is the length of an arc along the periphery of a disc. Fig. 6.1 shows a resulting model of a cell cytoplasm with $\sigma = 0.9 * 520$ nm and $\Lambda = \frac{2\pi r}{20}$, where r is the radius of the cell cytoplasm and we set $r = 5$ nm in this figure.

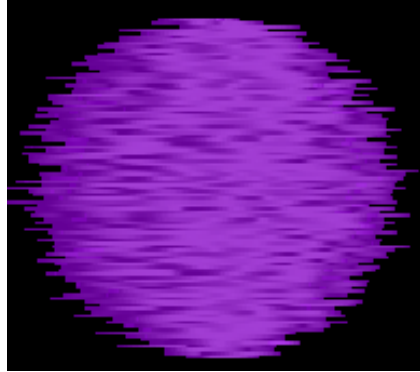


Figure 6.1: A cell cytoplasm model with the roughness model implemented. The following parameters are used: $\sigma = 0.9 * 520$ nm, $\Lambda = \frac{2\pi r}{20}$ and $r = 5$ nm, where r is the radius of the cell cytoplasm.

Cell models with two different (75 and 100) numbers of mitochondria having four different cell surface roughness amplitudes (1%, 3%, 5% and 7% of the cell diameter with a suitable correlation length of $2\pi * cell\ dia/100$) were used for numerical simulation in this study.

6.2.2 Numerical Simulation Models and the Angular Scattering Patterns

Lumerical FDTD [47], a popular commercial finite-difference time-domain (FDTD) solver, was utilized for solve the 3D full-vector Maxwell's equations. A plane-wave excitation source was used to interrogate the cell. Perfectly Matched Layer (PML) boundary conditions were employed along all the plane boundary walls of a cubic computational domain. A non-uniform custom meshing scheme with finer meshes in the vicinity of smaller length features (e.g., cell surface roughness) was used to maintain the accuracy of computation throughout the domain. The simulation was performed using three wavelengths (432 nm, 532 nm, and 632 nm). After solving the full-vectorial electric and magnetic fields computation, near-to-far field transformations were used to obtain the scattering patterns placed at different angular positions, resembling the placement of digital cameras that are expected to be used during the experiment process. In our study, scattering patterns were collected in the perfect-forward, perfect-backward, and side directions, where the center of the observation windows is placed at the 0, 180, and 90-degree angles with respect to the laser light, respectively. All the observation windows have an angular cone of 30 degrees. The light scattering from the whole cell or the large nuclei produces angular distribution corresponding to rings in the forward and backward directions and fringes in the side direction, which is similar to the Mie scattering patterns. Scattering on the small-scale structures, such as mitochondria, leads to the interference patterns displaying speckle distributions that dominate the far-field images in the side direction (row 4 in Fig. 2). We have also found that such small-scale variations in the cell surface roughness can also have a similar effect on the angular distribution of the scattered light. The backward scattered light (row 3 in Fig. 2) illustrates the superposition of the ring-like scattering patterns and the small-scale speckles due to the mitochondria. The forward scattered light (row 5 in Fig. 2) is dominated by the scattering from the cell and nucleus with the very faint contributions from the rest of the scattering centers.

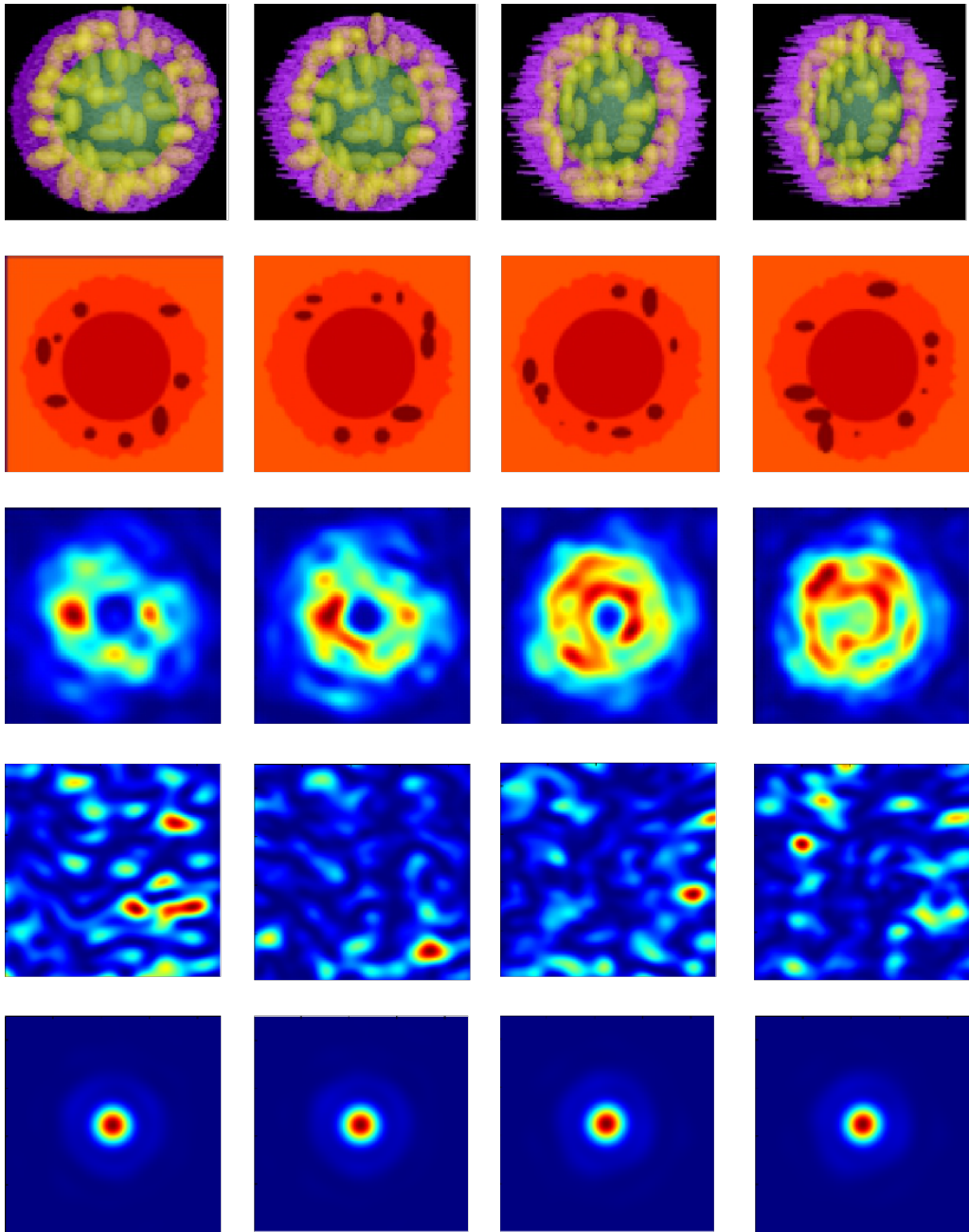


Figure 6.2: Example images of the cell models and their corresponding scattering patterns. The amplitudes of modulation for the cell surface roughness are 1%, 3%, 5% and 7% (sequentially from the left column to right column) of the the cell diameter (6.6 micron). Row 3-5 show the scattering patterns collected in the backward, side and forward direction. The wavelength is 432 nm.

6.3 Dataset for Machine Learning

The final dataset contains scattering patterns of 4 levels of roughness in the forward, side, and backward directions, with three wavelengths, 432 nm, 532 nm, and 632 nm. The number of scattering images for each sub-category is 100. During the experiments, 200 scattering patterns are used for both the simple and the complex cases to make a fair comparison.

Fig. 6.3, Fig. 6.4 and Fig. 6.5 display the simulated scattering patterns that are generated with different roughness value r , with wavelength 432 nm, 532 nm, 632 nm, in the forward, side, and backward direction, respectively. As shown in Fig. 6.3, the scattering pattern collected in the forward direction contains a bright circular spot centered in the image on a dark background. As the wavelength increases, the radius of the spot increases. Some circular fringe patterns can be observed along the radius of a spot (best shown in Fig. 1.2, row 5). There seems no visible difference when the roughness levels and the number of mitochondria are different. Fig. 6.4 displays the scattering patterns collected in the side direction, which are dominated by scattered speckles. The speckle size increases and the number of speckles decreases when the wavelength increases from 432 nm to 632 nm. Although the scattered patterns are different considering two different cells, there is no visible difference by observing the entire dataset when the roughness levels and the number of mitochondria are different. Fig. 6.5 shows the scattering patterns collected in the backward direction. As the roughness value increases, the speckles gradually form into bright circular fringes. Different visual patterns are observed when the wavelength varies from 432 nm to 632 nm. Especially, a bright spot appears in the center of images with the wavelength 632 nm. In contrast, a hollow is observed when the wavelength is 432 nm and the roughness level is 0.01, 0.03, 0.05. No apparent visual difference is observed when the number of mitochondria is different.

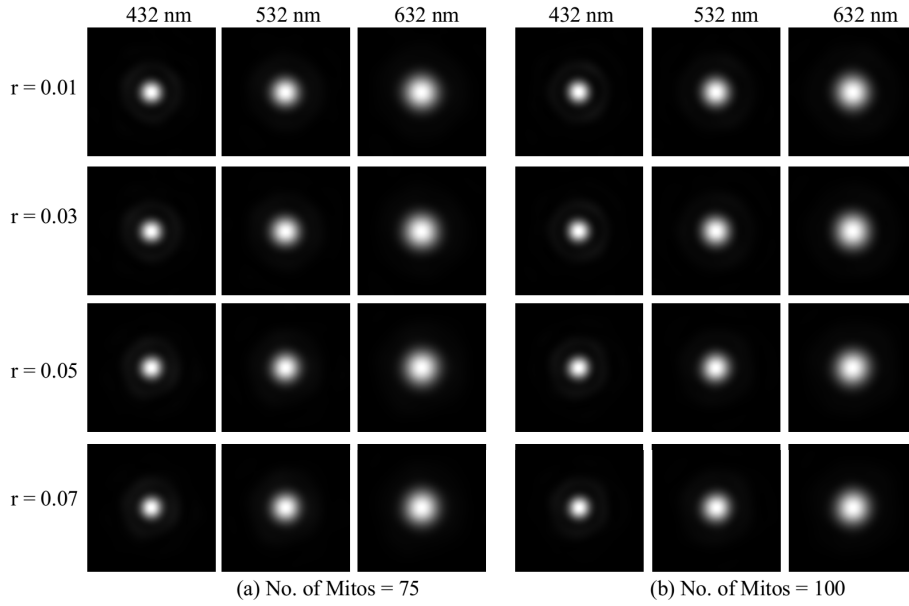


Figure 6.3: Example of simulated scattering images collected in the forward direction.

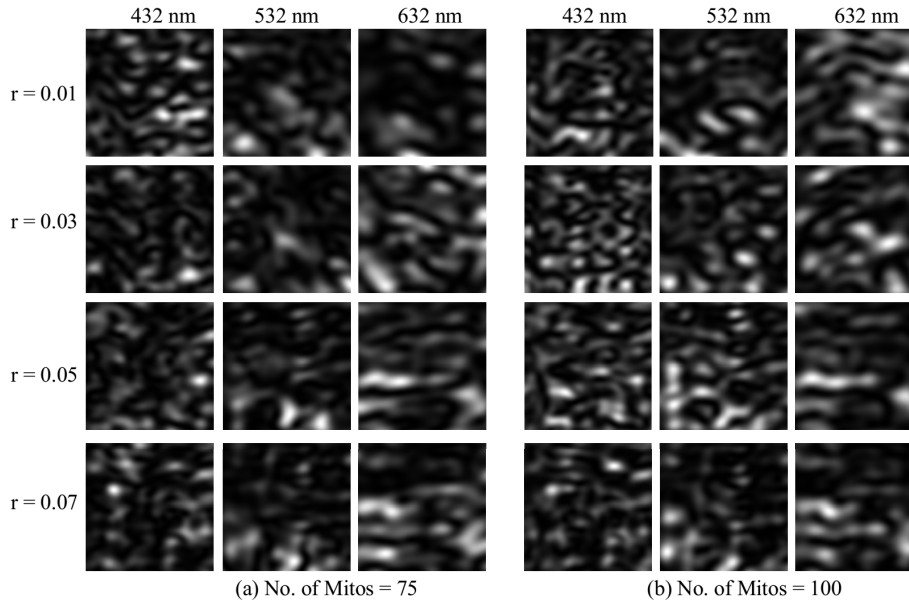


Figure 6.4: Example of simulated scattering images collected in the side direction.

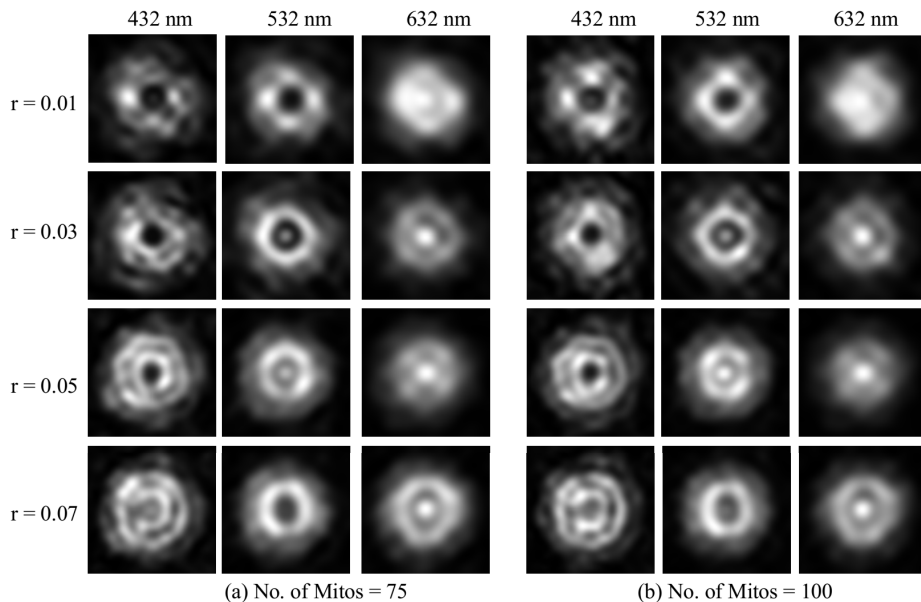


Figure 6.5: Example of simulated scattering images collected in the backward direction.

6.4 Experimental Results

6.4.1 Classification of Different Roughness Levels

In this section, scattering patterns with different roughness levels are analyzed, and their classification performances are also included. The surface roughness is indicated by r , and a larger value of r means a rougher cell surface. In a simple case, the roughness values of the scattering patterns are different ($r = 0.01$ and $r = 0.03$) while the number of mitochondria is the same (75 or 100). Experimental results regarding the simple cases are shown in Table 6.1 (the number of mitochondria is 75) and Table 6.2 (the number of mitochondria is 100). From Table 6.1 and Table 6.2, we can find that the scattered light patterns collected in all the three directions contain useful information regarding cell surface roughness since an over 72% of accuracy is observed for scattered light in the forward, side and backward directions. Scattering patterns collected in the backward direction perform the best, with almost 100% of accuracy, which indicates that backward scattering is most relevant to the cell surface roughness. This finding is reasonable as the majority of scattered

light in the backward direction may be from the cell surface, which is consistent with previous research that the backscattering is most relevant with the cell membrane [55]. The wavelength of the laser beam also plays an important role in the scattered light patterns. The forward scattered light with wavelength $\lambda = 632$ nm provides an accuracy of 98%, compared with around 76.0 when the wavelength is 432 nm or 532 nm.

Table 6.1: Classification of scattering patterns with $r = 0.01$ and $r = 0.03$. The mitochondria number is 75. 5-fold cross validation is used, and the average performance is given.

$\lambda = 432$ nm				
Direction	ACC	SEN	SPE	AUC
F	76.0	78.0	74.0	86.0
S	83.0	85.0	81.0	92.0
B	97.0	98.0	96.0	99.6
$\lambda = 532$ nm				
Direction	ACC	SEN	SPE	AUC
F	76.5	77.0	76.0	82.9
S	86.5	84.0	89.0	94.3
B	100	100	100	100
$\lambda = 632$ nm				
Direction	ACC	SEN	SPE	AUC
F	98	99	97	99.95
S	83	82	84	92.8
B	100	100	100	100

Table 6.2: Classification of scattering patterns with $r = 0.01$ and $r = 0.03$. The mitochondria number is 100. 5-fold cross validation is used, and the average performance is given.

$\lambda = 432 \text{ nm}$				
Direction	ACC	SEN	SPE	AUC
F	72.0	85.0	59.0	75.2
S	87.5	88.0	87.0	94.5
B	99	100	98	100
$\lambda = 532 \text{ nm}$				
Direction	ACC	SEN	SPE	AUC
F	72.5	70.0	75.0	78.9
S	85.5	87.0	84.0	93.5
B	100	100	100	100
$\lambda = 632 \text{ nm}$				
Direction	ACC	SEN	SPE	AUC
F	98.5	98.0	99.0	99.8
S	76.5	74.0	79.0	83.2
B	100	100	100	100

Table 6.3: Classification of scattering patterns with $r = 0.01$ and $r = 0.03$. For each group, an equal number of scattering patterns with 75 and 100 number of mitochondria are included. 5-fold cross validation is used, and the average performance is given.

$\lambda = 432 \text{ nm}$				
Direction	ACC	SEN	SPE	AUC
F	78.5	75.0	82.0	84.7
S	83.5	84.0	83.0	93.2
B	98.0	98.0	98.0	99.5
$\lambda = 532 \text{ nm}$				
Direction	ACC	SEN	SPE	AUC
F	75.5	77.0	74.0	83.3
S	85.5	86.0	85.0	93.3
B	100	100	100	100
$\lambda = 632 \text{ nm}$				
Direction	ACC	SEN	SPE	AUC
F	97.0	97.0	97.0	99.6
S	84.0	85.0	83.0	92.6
B	100	100	100	100

In a real case, the inner cellular structure is more complex, and there will be some randomness in the parameter space of cells. We simulate this com-

plex situation by mixing an equal number of scattering patterns with 75 and 100 mitochondria in each group. Experimental results are shown in Table 6.3, which are similar to the performance of the simple cases. We can still obtain an accuracy of 100% using scattering patterns in the backward direction. This indicates that the backward scattered light is not affected by the small variations (mitochondria number in this case) of the cell’s internal structure. Machine learning techniques can still learn useful information from the mixing samples.

In our previous experimental study of the NT and ST SH-SY5Y neuroblastoma cells (see chapter 5), we had found that the 2D light scattering patterns from the near direction (18 to 42 degrees) probing by a red laser (632 nm) provided the best results in distinguishing the NT and ST cells. The machine learning method achieved high performance with ACC = 91.11, SEN = 92.78, SPE = 89.44 and AUC = 96.95 (see Table 5.1). Since the non-treated and staurosporine-treated SH-SY5Y neuroblastoma cells have distinct cell membrane roughness (see Fig 5.1), we speculated that the difference in surface roughness could significantly contribute to their scattering patterns [85]. The results from Table 6.3 indicated the performance metrics for the forward (-15 to 15 degrees) and side (75 to 105 degrees) are in the range of 85 - 100 for identifying synthetic cells with different roughness levels. The results support the speculation suggested in [85]. In addition, the better performance of the forward direction versus side direction and red versus green laser also agree with the experimental data presented in chapter 5 [85].

6.4.2 Classification of Different Mitochondrial Numbers

In this section, we change the number of mitochondria and generate their simulated scattering patterns. The scattering patterns with different numbers of mitochondria are analyzed, and classification performances are also included. In the simple case, the mitochondrial number is different (75 or 100) while the roughness values are the same ($r = 0.01$ or $r = 0.03$). Experimental results are shown in Table 6.4 and Table 6.5. In complex cases, we set an equal number of scattering images with different roughness values ($r = 0.01$ and $r = 0.03$)

in each group.

From Table 6.4 and Table 6.5, we can see that the forward direction performs the best for laser light with different wavelengths, with an accuracy of over 86.5%. This indicates that the forward scattering carries the most useful information regarding the number of mitochondria. This may be due to the fact that a higher number of mitochondria may cause a high reflection index, which is most relevant with the forward scattered light [56]. The side direction always gets an accuracy of around 50%, which means it has failed in the discrimination of these two groups of cells. The side scattered light may not carry useful information regarding the mitochondria number. This is different from our previous research, as the side direction can carry useful information regarding the mitochondria distribution [78].

Table 6.4: Classification of scattering patterns with different numbers of mitochondria 75 and 100. The roughness value is $r = 0.01$. 5-fold cross validation is used, and the average performance is given.

$\lambda = 432 \text{ nm}$				
Direction	ACC	SEN	SPE	AUC
F	90.5	90.0	91.0	97.9
S	41.5	44.0	39.0	46.0
B	66.5	66 .0	67.0	75.6
$\lambda = 532 \text{ nm}$				
Direction	ACC	SEN	SPE	AUC
F	96.5	95.0	98.0	99.5
S	56.5	52.0	61.0	55.5
B	78.0	76.0	80.0	87.1
$\lambda = 632 \text{ nm}$				
Direction	ACC	SEN	SPE	AUC
F	98.5	100.0	97.0	99.9
S	65.0	63.0	67.0	70.8
B	66.0	80.0	52.0	71.2

Table 6.5: Classification of scattering patterns with different numbers of mitochondria 75 and 100. The roughness value is $r = 0.03$. 5-fold cross validation is used, and the average performance is given.

$\lambda = 432$ nm				
Direction	ACC	SEN	SPE	AUC
F	86.5	88.0	85.0	96.3
S	47.5	60.0	35.0	43.8
B	72.0	72.0	72.0	80.4
$\lambda = 532$ nm				
Direction	ACC	SEN	SPE	AUC
F	97.0	98.0	96.0	99.9
S	50.5	48.0	53.0	46.3
B	73.5	74.0	73.0	82.0
$\lambda = 632$ nm				
Direction	ACC	SEN	SPE	AUC
F	97	97	97	99.3
S	46.0	54.0	38.0	45.8
B	77.0	75.0	79.0	85.5

Experimental results of the complex case are shown in Table 6.6. We can still achieve similar performance when the wavelength are 532 nm and 632 nm.

Table 6.6: Classification of scattering patterns with different numbers of mitochondria 75 and 100. For each group, an equal number of scattering patterns with $r = 0.01$ and $r = 0.03$ roughness values are included. 5-fold cross validation is used, and the average performance is given.

$\lambda = 432$ nm				
Direction	ACC	SEN	SPE	AUC
F	91.0	89.0	93.0	97.5
S	45.0	52.0	38.0	44.7
B	68.5	70.0	67.0	75.9
$\lambda = 532$ nm				
Direction	ACC	SEN	SPE	AUC
F	97.5	98	97	99.3
S	56.0	57.0	55.0	54.7
B	77.5	79.0	76.0	86.2
$\lambda = 632$ nm				
Direction	ACC	SEN	SPE	AUC
F	97.5	96.0	99.0	100
S	61.5	66.0	57.0	64.6
B	69.5	60.0	79.0	76.7

6.5 Discussions

6.5.1 Quantitative Result Analysis

As shown in Sec. 6.4.1, the backward scattered light is most useful for the classification of the surface roughness. To analyze this phenomenon, some example images are shown in Fig. 6.6. As we can see from Fig. 6.6, there are distinct visual differences when the roughness varies. A centered bright spot and a circular structure are observed as the roughness increases. Therefore, excellent performance can be achieved by using scattering images collected in the backward direction.

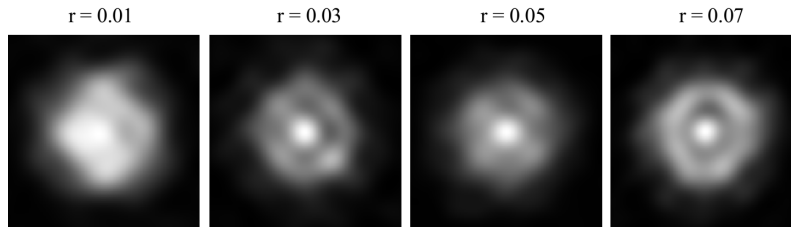


Figure 6.6: Scattering images with different roughness values r , which are collected in the backward direction. The mitochondria number is 100.

Similarly, Sec. 6.4.1 has shown that the forward direction is most useful for the classification of the mitochondrial number. The original scattering patterns in the forward direction have been shown in Fig. 6.3, where no visible difference is observed given cells with different numbers of mitochondria. This may be due to the fact that the original images are represented in linear scale, and it is too dark to visualize the fine textures. Hence, we first preprocess the original images using log transform. The processed images are shown in Fig. 6.7. As we can see from Fig. 6.7, as the mitochondrial number decreases, the circular fringe patterns shrink and become deformed, and more textures are found on the boundary of the scattering patterns.

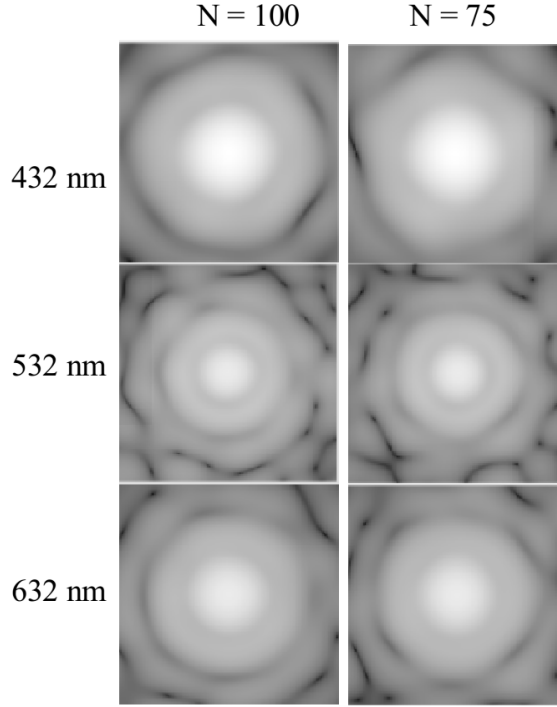


Figure 6.7: Example images with different mitochondria numbers N . The roughness value r is 0.01.

6.5.2 t-SNE Visualization

The t-SNE [49] allows us to visualize the distribution of high dimensional data to some degree. The t-SNE technique first map the original high-dimensional features from 1920 to 50 dimension by using PCA, and then to 2 by using the t-SNE Barnes-Hut algorithm [49]. The x-axis and y-axis correspond to the values of the two features after using the t-SNE method.

The t-SNE plots of light scattering patterns collected in the forward, side, and backward directions with different roughness levels are shown in Fig. 6.8. The number of mitochondria is $N = 75$. As we can see from Fig. 6.8, the two groups of data (with different levels of roughness) are well separated for light scattering patterns collected in the backward direction with different wavelengths. In addition, clustering behavior is also observed in the forward direction when the wavelength $\lambda = 632$ nm. This means that these light scattering patterns have distinct features in the raw feature space, which matches with our previous findings in Tables 6.3, 6.4, 6.5. Notice that we can also ob-

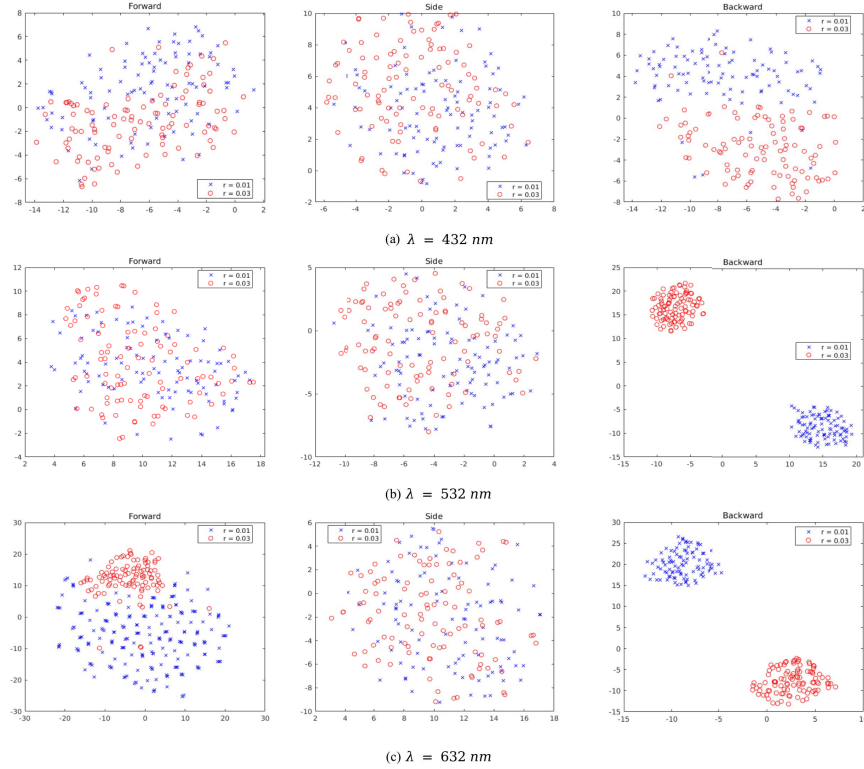


Figure 6.8: t-SNE visualization of light scattering patterns collected in the forward, side and backward directions with $\lambda = 432$ nm, 532 nm and 632 nm, and $N = 75$. The two groups of data with different roughness values r are denoted using different colours.

serve strong appearance variations from the original light scattering patterns collected in the backward direction when the roughness levels are different, which has been shown in Fig. 6.6.

We also provide the t-SNE scatter plot of light scattering patterns collected in the forward, side, and backward with $\lambda = 432$ nm, 532 nm, and 632 nm in Fig. 6.9. The roughness value is $r = 0.01$. The two groups of data with different numbers of mitochondria N are denoted using different colors. As Fig. 6.9 shows, the majority of the two groups of data from the forward direction can be well separated, and only a small amount of points are overlapped and mixed. This indicates that the forward scattered light carries useful information regarding the number of mitochondria which is consistent with the results shown in Table 6.4, Table 6.5 and Table 6.6.

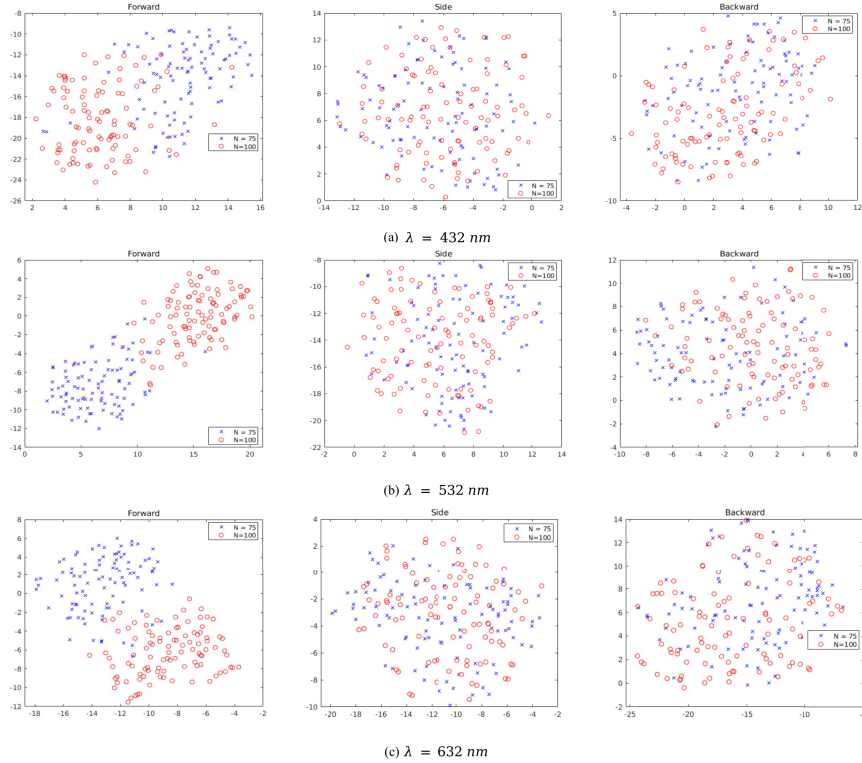


Figure 6.9: t-SNE visualization of scattering patterns collected in the forward, side and backward direction with wavelength $\lambda = 432$ nm, 532 nm and 632 nm. The roughness value is $r = 0.01$. The two groups of data with different mitochondrial number N are denoted using different colours.

6.5.3 Insights Relevant to Previous Studies

In the previous chapter, we obtained angular light scattering patterns in the side ($79^\circ - 101^\circ$) and near-forward direction ($18^\circ - 42^\circ$) probing by green (532 nm) or red (633 nm) laser light. We observed that the near-forward scattering using red light provided the best results in discriminating between the non-treated and staurosporine-treated cells. Satisfactory performance with ACC = 91%, SEN = 92%, SPE = 89% and AUC = 97% [85] were achieved in discriminating the two groups of cells by using our ML analysis method. The NT and ST SH-SY5Y neuroblastoma cells have significant differences in cell surface roughness, as observed from their SEM images [85]. We proposed that the differences in the surface roughness contributed to their unique scattering patterns, which can be effectively captured by the machine learning analysis [85]. The results from Table 6.3 shows that the performance metric for for-

ward and side scattering using red laser light are in the range of [83, 100] for identifying cells with different roughness, which are consistent with the results presented in [85]. In addition, the better performance of forward direction versus side direction and red versus green laser light is also consistent with the experimental results presented in [85]. While the performance of discriminating the two groups of cells in the near forward direction was better than the side direction in [85], the side direction performance was reasonable. Results from section 6.4.2 indicate poor performance for side direction in discriminating the variations in the number of mitochondria in cells. These results suggest that the change in the number of mitochondria in the treated cell as a result of apoptosis induced by staurosporine [26] might not contribute significantly to the differences in the angular light scattering patterns of the treated and non-treated cells in [85].

It was well-established that light scattering in the side direction can provide information about small internal organelles such as mitochondria [86], [87]. Several studies [71], [76]–[79] reported that single-cell light scattering patterns depended on mitochondrial distributions and formed the basis for the identification and analysis of many cells. However, as shown in Tables 4–6, the performance of discrimination of variations in the number of mitochondria in cells from their light scattering patterns collected in the side direction is poor. Although the side scattered light carries valuable information about the cells’ mitochondrial distributions (distinct scattering patterns are observed with the aggregated and the randomly distributed mitochondria) [78], it may not carry helpful information regarding the number of mitochondria.

6.6 Conclusions

In this study, multi-wavelength multi-direction laser scattering patterns have been achieved with numerical simulation. Two factors, including the cell surface roughness and the number of mitochondria, have been analyzed for their roles in contributing to specific scattering patterns. Machine learning techniques have been used to analyze and classify these scattering patterns. Re-

sults of the analysis have shown that both roughness values and mitochondria number affect scattering patterns. Especially, the backward light scattering is the best direction for characterizing the cell surface roughness, while the forward light scattering is the best direction for characterizing the mitochondria number. The results in this chapter are consistent with the experimental results in chapter 5 and provide insights into the origins of the experimental angular scattering patterns. This research provides a theoretical analysis of the contributions of cell surface roughness and mitochondria number in scattering patterns, which is potentially useful for the statistical analysis and measurements of inner-cellular organelles. A systematic study about the influences of the wavelength and angular direction has been included, which is beneficial for the real experiment setup of future research.

Chapter 7

Conclusions and Future Research Directions

In this thesis, machine learning methods, especially the CNN based methods, have been proposed for the automatic analysis of dermoscopic images and laser scattering images. In this chapter, we summarize the main contributions of each chapter and discuss potential future research work.

7.1 Main Contributions

Chapter 3 presents the automatic system for accurate skin lesion segmentation. Compared with the multi-task learning methods, the proposed method uses edge prediction as an auxiliary task, which does not require extra labeling efforts. The proposed model makes predictions of the boundaries and the segmentation mask by learning two sub-networks simultaneously, with cross-connected intermediate feature maps. Experimental results prove that the auxiliary task framework can promote the skin lesion segmentation task. State-of-the-art performance has been obtained.

Chapter 4 presents a novel technique for skin lesion classification based on mid-level feature learning. The learned features are more discriminative and representative compared to the original features by using distance metric learning. The features can capture the relationships between samples and hence are more robust. Experimental results have verified the effectiveness of the mid-level features.

Chapter 5 presents a technique for label-free cell classification using laser scattering patterns. A pretrained CNN model is used as an off-the-shelf feature extractor. Results show that the proposed technique can successfully discriminate between the two groups of cells: the staurosporine-treated and non-treated SH-SY5Y neuroblastoma cells, which is beneficial for a better understanding of the PD disease. Superior performance has been achieved compared with conventional machine learning methods.

Chapter 6 presents a systematic and thorough study towards the influences of two cell properties (surface roughness and mitochondria number) in contributing to the scattering patterns. Multi-wavelength multi-direction laser scattering images of the same single cell have been achieved by simulation. Analysis of these simulated laser scattering images has shown that both roughness values and mitochondria number affect scattering patterns. Especially, the backward light scattering is the best direction for characterizing the surface roughness, while the forward light scattering is the best direction for characterizing the mitochondria number. This research provides a theoretical analysis of the contributions of cell surface roughness and mitochondria number in scattering patterns, which is potentially useful for the statistical analysis and measurements of inner-cellular organelles.

7.2 Future Research Directions

- Skin Lesion Analysis with Vision Transformer

The CNN methods do not encode information about the relative position of local features. Recently, some methods based on vision transformers emerged in various computer vision tasks. The vision transformers can model the long-range dependencies of the input image patches. Since the pigment regions of the skin lesions and the surrounding skin areas share strong visual dependencies, the introduction of the transformer models may promote the segmentation and classification performance.

- **Laser Scattering Images with Hybrid CNN**

Existing deep learning methods for laser scattering images use popular CNN models for cell classification. The scattered speckles of laser scatter patterns may share dependencies, which have been ignored in literature works. A hybrid CNN which integrates the dependency information of local patches (e.g., by using the adjacency matrix of the feature maps) can provide more information regarding the relationships of local patches, which may facilitate the training of the CNN-based methods.

References

- [1] Q. Abbas, I. Fondón, A. Sarmiento, and M. E. Celebi, “An improved segmentation method for non-melanoma skin lesions using active contour model,” in *International Conference Image Analysis and Recognition*, 2014, pp. 193–200.
- [2] K. Abhishek, G. Hamarneh, and M. S. Drew, “Illumination-based transformations improve skin lesion segmentation in dermoscopic images,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2020, pp. 728–729.
- [3] T. Ahonen, A. Hadid, and M. Pietikainen, “Face description with local binary patterns: Application to face recognition,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 12, pp. 2037–2041, 2006.
- [4] A.-R. Ali, M. S. Couceiro, and A. E. Hassenian, “Melanoma detection using fuzzy c-means clustering coupled with mathematical morphology,” in *International Conference on Hybrid Intelligent Systems (HIS)*, 2014, pp. 73–78.
- [5] T. Ando, R. Horisaki, and J. Tanida, “Speckle-learning-based object recognition through scattering media,” *Optics Express*, vol. 23, no. 26, pp. 33 902–33 910, 2015.
- [6] G. Argenziano, G. Fabbrocini, P. Carli, V. De Giorgi, E. Sammarco, and M. Delfino, “Epiluminescence microscopy for the diagnosis of doubtful melanocytic skin lesions: Comparison of the abcd rule of dermatoscopy and a new 7-point checklist based on pattern analysis,” *Archives of Dermatology*, vol. 134, no. 12, pp. 1563–1570, 1998.
- [7] V. Badrinarayanan, A. Kendall, and R. Cipolla, “Segnet: A deep convolutional encoder-decoder architecture for image segmentation,” *arXiv preprint arXiv:1511.00561*, 2015.
- [8] S. Ballaz, I. Morales, M. Rodríguez, and J. A. Obeso, “Ascorbate prevents cell death from prolonged exposure to glutamate in an in vitro model of human dopaminergic neurons,” *Journal of Neuroscience Research*, vol. 91, no. 12, pp. 1609–1617, 2013.
- [9] M. Berseth, “Isic 2017-skin lesion analysis towards melanoma detection,” *arXiv preprint arXiv:1703.00523*, 2017.

- [10] L. Bi, J. Kim, E. Ahn, and D. Feng, “Automatic skin lesion analysis using large-scale dermoscopy images and deep residual networks,” *arXiv preprint arXiv:1703.04197*, 2017.
- [11] L. Bi, J. Kim, E. Ahn, A. Kumar, D. Feng, and M. Fulham, “Step-wise integration of deep class-specific learning for dermoscopic image segmentation,” *Pattern Recognition*, vol. 85, pp. 78–89, 2019.
- [12] A. Bissoto, M. Fornaciali, E. Valle, and S. Avila, “(de) constructing bias on skin lesion datasets,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2019.
- [13] D. Borys, P. Kowalska, M. Frackiewicz, and Z. Ostrowski, “A simple hair removal algorithm from dermoscopic images,” in *International Conference on Bioinformatics and Biomedical Engineering*, 2015, pp. 262–273.
- [14] H. Chang, “Skin cancer reorganization and classification with deep neural network,” *arXiv preprint arXiv:1703.00534*, 2017.
- [15] M. Clinic, “Squamous cell carcinoma of the skin,” *Online*, 2021. [Online]. Available: <https://www.mayoclinic.org/diseases-conditions/squamous-cell-carcinoma/symptoms-causes/syc-20352480>.
- [16] N. C. Codella, D. Gutman, M. E. Celebi, and et.al., “Skin lesion analysis toward melanoma detection: A challenge at the 2017 international symposium on biomedical imaging,” in *International Symposium on Biomedical Imaging*, 2018, pp. 168–172.
- [17] N. Dalal and B. Triggs, “Histograms of oriented gradients for human detection,” in *2005 IEEE Conference on Computer Vision and Pattern Recognition*, vol. 1, 2005, pp. 886–893.
- [18] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, “Imagenet: A large-scale hierarchical image database,” in *2009 IEEE Conference on Computer Vision and Pattern Recognition*, 2009, pp. 248–255.
- [19] M. Ebner, “Combining white-patch retinex and the gray world assumption to achieve color constancy for multiple illuminants,” in *Joint Pattern Recognition Symposium*, 2003, pp. 60–67.
- [20] —, *Color constancy*. John Wiley & Sons, 2007, vol. 7.
- [21] A. Esteva, B. Kuprel, R. A. Novoa, J. Ko, S. M. Swetter, H. M. Blau, and S. Thrun, “Dermatologist-level classification of skin cancer with deep neural networks,” *Nature*, vol. 542, no. 7639, pp. 115–118, 2017.
- [22] Z. Ge, S. Demyanov, B. Bozorgtabar, and et.al., “Exploiting local and generic features for accurate skin lesions classification using clinical and dermoscopy imaging,” in *International Symposium on Biomedical Imaging*, 2017, pp. 986–990.

- [23] Z. Ge, S. Demyanov, R. Chakravorty, A. Bowling, and R. Garnavi, “Skin disease recognition using deep saliency features and multimodal learning of dermoscopy and clinical images,” in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 2017, pp. 250–258.
- [24] I. González-Díaz, “Dermaknet: Incorporating the knowledge of dermatologists to convolutional neural networks for skin lesion diagnosis,” *IEEE Journal of Biomedical and Health Informatics*, vol. 23, no. 2, pp. 547–559, 2019.
- [25] M. Goyal, A. Oakley, P. Bansal, D. Dancey, and M. H. Yap, “Skin lesion segmentation in dermoscopic images with ensemble deep learning methods,” *IEEE Access*, vol. 8, pp. 4171–4181, 2019.
- [26] D. R. Green and J. C. Reed, “Mitochondria and apoptosis,” *Science*, vol. 281, no. 5381, pp. 1309–1312, 1998.
- [27] Y. Gu, J. Zhou, and B. Qian, “Melanoma detection based on mahalanobis distance learning and constrained graph regularized nonnegative matrix factorization,” in *IEEE Winter Conference on Applications of Computer Vision (WACV)*, 2017, pp. 797–805.
- [28] B. P. Hazen, A. C. Bhatia, T. Zaim, and R. T. Brodell, “The clinical diagnosis of early malignant melanoma: expansion of the ABCD criteria to improve diagnostic sensitivity,” *Dermatology Online Journal*, vol. 5, no. 2, 1999.
- [29] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in *IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 770–778.
- [30] K. Hiramatsu, T. Ideguchi, Y. Yonamine, S. Lee, Y. Luo, K. Hashimoto, T. Ito, M. Hase, J.-W. Park, Y. Kasai, *et al.*, “High-throughput label-free molecular fingerprinting flow cytometry,” *Science Advances*, vol. 5, no. 1, 2019.
- [31] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, “Densely connected convolutional networks,” in *IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 4700–4708.
- [32] J. Humayun, A. S. Malik, and N. Kamel, “Multilevel thresholding for segmentation of pigmented skin lesions,” in *IEEE International Conference on Imaging Systems and Techniques*, 2011, pp. 310–314.
- [33] S. Ioffe and C. Szegedy, “Batch normalization: Accelerating deep network training by reducing internal covariate shift,” *arXiv preprint arXiv: 1502.03167*, 2015.

- [34] M. H. Jafari, N. Karimi, E. Nasr-Esfahani, S. Samavi, S. M. R. Soroushmehr, K. Ward, and K. Najarian, “Skin lesion segmentation in clinical images using deep learning,” in *International Conference on Pattern Recognition (ICPR)*, 2016, pp. 337–342.
- [35] M. H. Jafari, S. Samavi, S. M. R. Soroushmehr, H. Mohaghegh, N. Karimi, and K. Najarian, “Set of descriptors for skin cancer diagnosis using non-dermoscopic color images,” in *International Conference on Image Processing (ICIP)*, 2016, pp. 2638–2642.
- [36] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “Imagenet classification with deep convolutional neural networks,” *Advances in Neural Information Processing Systems*, vol. 25, 2012.
- [37] T. Lee, V. Ng, R. Gallagher, A. Coldman, and D. McLean, “Dullrazor®: A software approach to hair removal from images,” *Computers in Biology and Medicine*, vol. 27, no. 6, pp. 533–543, 1997.
- [38] Y. Li and L. Shen, “Skin lesion analysis towards melanoma detection using deep learning network,” *Sensors*, vol. 18, no. 2, 2018.
- [39] S. Liao and S. Z. Li, “Efficient psd constrained asymmetric metric learning for person re-identification,” in *IEEE International Conference on Computer Vision*, 2015, pp. 3685–3693.
- [40] C. Liu, C. E. Capjack, and W. Rozmus, “3-D simulation of light scattering from biological cells and cell differentiation,” *Journal of Biomedical Optics*, vol. 10, no. 1, 2005.
- [41] J. Liu, Y. Xu, W. Wang, Y. Wen, H. Hong, J. Q. Lu, P. Tian, and X.-H. Hu, “Machine learning of diffraction image patterns for accurate classification of cells modeled with different nuclear sizes,” *Journal of Biophotonics*, vol. 13, no. 9, 2020.
- [42] L. Liu, X. Lu, Y. Yuan, and X. Li, “Person re-identification by bidirectional projection,” in *Proceedings of International Conference on Internet Multimedia Computing and Service*, ACM, 2014, pp. 1–5.
- [43] L. Liu, L. Mou, X. X. Zhu, and M. Mandal, “Skin lesion segmentation based on improved u-net,” in *2019 IEEE Canadian Conference of Electrical and Computer Engineering (CCECE)*, 2019, pp. 1–4.
- [44] —, “Automatic skin lesion classification based on mid-level feature learning,” *Computerized Medical Imaging and Graphics*, vol. 84, 2020.
- [45] S. Liu, Z. Yuan, X. Qiao, Q. Liu, K. Song, B. Kong, and X. Su, “Light scattering pattern specific convolutional network static cytometry for label-free classification of cervical cells,” *Cytometry Part A*, 2021.
- [46] D. G. Lowe, “Object recognition from local scale-invariant features,” in *IEEE International Conference on Computer Vision*, vol. 2, 1999, pp. 1150–1157.

- [47] “Lumerical FDTD solutions software,” [Online]. Available: <https://www.lumerical.com/>.
- [48] Z. Ma and J. M. R. Tavares, “Effective features to classify skin lesions in dermoscopic images,” *Expert Systems with Applications*, vol. 84, pp. 92–101, 2017.
- [49] L. v. d. Maaten and G. Hinton, “Visualizing data using t-SNE,” *Journal of Machine Learning Research*, vol. 9, no. 11, pp. 2579–2605, 2008.
- [50] A. Mahbod, G. Schaefer, I. Ellinger, R. Ecker, A. Pitiot, and C. Wang, “Fusing fine-tuned deep features for skin lesion classification,” *Computerized Medical Imaging and Graphics*, vol. 71, pp. 19–29, 2019.
- [51] K.-K. Maninis, S. Caelles, J. Pont-Tuset, and L. Van Gool, “Deep extreme cut: From extreme points to object segmentation,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 616–625.
- [52] M. A. Al-Masni, M. A. Al-Antari, M.-T. Choi, S.-M. Han, and T.-S. Kim, “Skin lesion segmentation in dermoscopy images via deep full resolution convolutional networks,” *Computer Methods and Programs in Biomedicine*, vol. 162, pp. 221–231, 2018.
- [53] K. Matsunaga, A. Hamada, A. Minagawa, and H. Koga, “Image classification of melanoma, nevus and seborrheic keratosis by deep neural network ensemble deep convolutional encoder-decoder architecture for image segmentation,” *arXiv preprint arXiv:1703.03108*, 2017.
- [54] A. Menegola, J. Tavares, M. Fornaciali, L. T. Li, S. Avila, and E. Valle, “RECOD titans at ISIC challenge 2017,” *arXiv preprint arXiv:1703.04819*, 2017.
- [55] R. A. Meyer, “Light scattering from biological cells: Dependence of backscatter radiation on membrane thickness and refractive index,” *Applied Optics*, vol. 18, no. 5, pp. 585–588, 1979.
- [56] P. Mullaney, M. Van Dilla, J. Coulter, and P. Dean, “Cell sizing: A light scattering photometer for rapid volume determination,” *Review of Scientific Instruments*, vol. 40, no. 8, pp. 1029–1032, 1969.
- [57] M. M. Muqit, S. M. Davidson, M. D. Payne Smith, L. P. MacCormac, S. Kahns, P. H. Jensen, N. W. Wood, and D. S. Latchman, “Parkin is recruited into aggresomes in a stress-specific manner: Over-expression of parkin reduces aggresome formation but can be dissociated from parkin’s effect on neuronal survival,” *Human Molecular Genetics*, vol. 13, no. 1, pp. 117–135, 2004.
- [58] A. Oakley, “Dermoscopy,” *Online*, [Online]. Available: <https://dermnetz.org/topics/dermoscopy/>.

- [59] R. B. Oliveira, J. P. Papa, A. S. Pereira, and J. M. R. Tavares, “Computational methods for pigmented skin lesion classification in images: Review and future trends,” *Neural Computing and Applications*, vol. 29, no. 3, pp. 613–636, 2018.
- [60] Y.-L. Pan, M. J. Berg, S. S.-M. Zhang, H. Noh, H. Cao, R. K. Chang, and G. Videen, “Measurement and autocorrelation analysis of two-dimensional light-scattering patterns from living cells for label-free classification,” *Cytometry Part A*, vol. 79, no. 4, pp. 284–292, 2011.
- [61] H. Pehamberger, A. Steiner, and K. Wolff, “In vivo epiluminescence microscopy of pigmented skin lesions. i. pattern analysis of pigmented skin lesions,” *Journal of the American Academy of Dermatology*, vol. 17, no. 4, pp. 571–583, 1987.
- [62] J. Picot, C. L. Guerin, C. Le Van Kim, and C. M. Boulanger, “Flow cytometry: Retrospective, fundamentals and recent instrumentation,” *Cytotechnology*, vol. 64, no. 2, pp. 109–130, 2012.
- [63] M. Ratajczak, M. Kucia, and J. Ratajczak, *Very small embryonic-like (vsel) stem cells and methods of isolating and using the same cross reference to related applications*, 2007.
- [64] P. P. Rebouças Filho, S. A. Peixoto, R. V. M. da Nóbrega, D. J. Hemanth, A. G. Medeiros, A. K. Sangaiah, and V. H. C. de Albuquerque, “Automatic histologically-closer classification of skin lesions,” *Computerized Medical Imaging and Graphics*, vol. 68, pp. 40–54, 2018.
- [65] F. Riaz, S. Naeem, R. Nawaz, and M. Coimbra, “Active contours based segmentation and lesion periphery analysis for characterization of skin lesions in dermoscopy images,” *IEEE Journal of Biomedical and Health Informatics*, vol. 23, no. 2, pp. 489–500, 2018.
- [66] V. Ribeiro, S. Avila, and E. Valle, “Handling inter-annotator agreement for automated skin lesion segmentation,” *arXiv preprint arXiv:1906.02415*, 2019.
- [67] O. Ronneberger, P. Fischer, and T. Brox, “U-net: Convolutional networks for biomedical image segmentation,” in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 2015, pp. 234–241.
- [68] G. I. Ruban, V. V. Berdnik, D. V. Marinitch, N. V. Goncharova, and V. A. Loiko, “Light scattering and morphology of the lymphocyte as applied to flow cytometry for distinguishing healthy and infected individuals,” *Journal of Biomedical Optics*, vol. 15, no. 5, 2010.
- [69] M. M. K. Sarker, H. A. Rashwan, F. Akram, S. F. Banu, A. Saleh, V. K. Singh, F. U. Chowdhury, S. Abdulwahab, S. Romani, P. Radeva, *et al.*, “SLSDeep: Skin lesion segmentation based on dilated residual and pyramid pooling networks,” in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 2018, pp. 21–29.

- [70] S. Schaefer, T. McPhail, and J. Warren, “Image deformation using moving least squares,” in *ACM transactions on graphics (TOG)*, vol. 25, 2006, pp. 533–540.
- [71] H. Shahin, M. Gupta, A. Janowska-Wieczorek, W. Rozmus, and Y. Y. Tsui, “Physical characterization of hematopoietic stem cells using multidirectional label-free light scatterings,” *Optics Express*, vol. 24, no. 25, pp. 28 877–28 888, 2016.
- [72] R. L. Siegel, K. D. Miller, and A. Jemal, “Cancer statistics, 2019,” *CA: a cancer journal for clinicians*, vol. 69, no. 1, pp. 7–34, 2019.
- [73] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” *arXiv preprint arXiv:1409.1556*, 2014.
- [74] V. K. Singh, M. Abdel-Nasser, H. A. Rashwan, F. Akram, N. Pandey, A. Lalande, B. Presles, S. Romani, and D. Puig, “FCA-Net: Adversarial learning for skin lesion segmentation based on multi-scale features and factorized channel attention,” *IEEE Access*, vol. 7, pp. 130 552–130 565, 2019.
- [75] W. Stolz, A. Riemann, A. Cognetta, L. Pillet, W. Abmayr, D. Holzel, P. Bilek, F. Nachbar, and M. Landthaler, “ABCD rule of dermatoscopy—a new practical method for early recognition of malignant-melanoma,” *European Journal of Dermatology*, vol. 4, no. 7, pp. 521–527, 1994.
- [76] X.-T. Su, C. Capjack, W. Rozmus, and C. Backhouse, “2D light scattering patterns of mitochondria in single cells,” *Optics Express*, vol. 15, no. 17, pp. 10 562–10 575, 2007.
- [77] X.-T. Su, K. Singh, W. Rozmus, C. Backhouse, and C. Capjack, “Light scattering characterization of mitochondrial aggregation in single cells,” *Optics Express*, vol. 17, no. 16, pp. 13 381–13 388, 2009.
- [78] X. Su, M. Gupta, C. E. Capjack, Y. Y. Tsui, Y. Qiu, L. Marquez-Curtis, A. Janowska-Wieczorek, and W. Rozmus, “Label-free and noninvasive optical detection of the distribution of nanometer-size mitochondria in single cells,” *Journal of Biomedical Optics*, vol. 16, no. 6, 2011.
- [79] X. Su, S. E. Kirkwood, M. Gupta, L. Marquez-Curtis, Y. Qiu, A. Janowska-Wieczorek, W. Rozmus, and Y. Y. Tsui, “Microscope-based label-free microfluidic cytometry,” *Optics Express*, vol. 19, no. 1, pp. 387–398, 2011.
- [80] X. Su, T. Yuan, Z. Wang, K. Song, R. Li, C. Yuan, and B. Kong, “Two-dimensional light scattering anisotropy cytometry for label-free classification of ovarian cancer cells via machine learning,” *Cytometry Part A*, vol. 97, no. 1, pp. 24–30, 2020.

- [81] Y. Suzuki, K. Kobayashi, Y. Wakisaka, D. Deng, S. Tanaka, C.-J. Huang, C. Lei, C.-W. Sun, H. Liu, Y. Fujiwaki, *et al.*, “Label-free chemical imaging flow cytometry by high-speed multicolor stimulated raman scattering,” *Proceedings of the National Academy of Sciences*, vol. 116, no. 32, pp. 15 842–15 848, 2019.
- [82] J. Tang, “A multi-direction gvf snake for the segmentation of skin cancer images,” *Pattern Recognition*, vol. 42, no. 6, pp. 1172–1179, 2009.
- [83] M. T. B. Toossi, H. R. Pourreza, H. Zare, M.-H. Sigari, P. Layegh, and A. Azimi, “An effective hair removal algorithm for dermoscopy images,” *Skin Research and Technology*, vol. 19, no. 3, pp. 230–235, 2013.
- [84] W. Tu, X. Liu, W. Hu, and Z. Pan, “Dense-residual network with adversarial learning for skin lesion segmentation,” *IEEE Access*, vol. 7, pp. 77 037–77 051, 2019.
- [85] W. Y. Wan, L. Liu, X. Liu, W. Wang, M. Z. Islam, C. Dong, C. R. Garen, M. T. Woodside, M. Gupta, M. Mandal, *et al.*, “Integration of light scattering with machine learning for label free cell detection,” *Biomedical Optics Express*, vol. 12, no. 6, pp. 3512–3529, 2021.
- [86] J. J. Wang, L. Han, Y. P. Han, G. Gouesbet, X. Wu, and Y. Wu, “Shaped beam scattering from a single lymphocyte cell by generalized lorenz–mie theory,” *Journal of Quantitative Spectroscopy and Radiative Transfer*, vol. 133, pp. 72–80, 2014.
- [87] D. Watson, N. Hagen, J. Diver, P. Marchand, and M. Chachisvilis, “Elastic light scattering from single cells: Orientational dynamics in optical trap,” *Biophysical Journal*, vol. 87, no. 2, pp. 1298–1306, 2004.
- [88] Z. Wei, H. Song, L. Chen, Q. Li, and G. Han, “Attention-based denseunet network with adversarial training for skin lesion segmentation,” *IEEE Access*, vol. 7, pp. 136 616–136 629, 2019.
- [89] A. Wong, J. Scharcanski, and P. Fieguth, “Automatic skin lesion segmentation via iterative stochastic region merging,” *IEEE Transactions on Information Technology in Biomedicine*, vol. 15, no. 6, pp. 929–936, 2011.
- [90] Y. Xie, J. Zhang, Y. Xia, and C. Shen, “A mutual bootstrapping model for automated skin lesion segmentation and classification,” *IEEE Transactions on Medical Imaging*, vol. 39, no. 7, pp. 2482–2493, 2020.
- [91] X. Yang, Z. Zeng, S. Y. Yeo, C. Tan, H. L. Tey, and Y. Su, “A novel multi-task deep learning model for skin lesion segmentation and classification,” *arXiv preprint arXiv:1703.01025*, 2017.
- [92] F. Yu and V. Koltun, “Multi-scale context aggregation by dilated convolutions,” *arXiv preprint arXiv:1511.07122*, 2015.
- [93] Y. Yuan, “Automatic skin lesion segmentation with fully convolutional-deconvolutional networks,” *arXiv preprint arXiv:1703.05165*, 2017.

- [94] Y. Yuan, M. Chao, and Y.-C. Lo, “Automatic skin lesion segmentation using deep fully convolutional networks with jaccard distance,” *IEEE Transactions on Medical Imaging*, vol. 36, no. 9, pp. 1876–1886, 2017.
- [95] L. Zhang, Y. Qin, K.-X. Li, X. Zhao, Y.-F. Xing, H. Zhao, Z. Jiang, W. Chen, N.-L. Yao, and L. Yuan, “Light scattering properties in spatial planes for label free cells with different internal structures,” *Optical and Quantum Electronics*, vol. 47, no. 5, pp. 1005–1025, 2015.
- [96] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia, “Pyramid scene parsing network,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 2881–2890.
- [97] F. Zhuang, Z. Qi, K. Duan, D. Xi, Y. Zhu, H. Zhu, H. Xiong, and Q. He, “A comprehensive survey on transfer learning,” *Proceedings of the IEEE*, vol. 109, no. 1, pp. 43–76, 2020.
- [98] E. K. Zuba-Surma, I. Klich, N. Greco, M. J. Laughlin, J. Ratajczak, and M. Z. Ratajczak, “Optimization of isolation and further characterization of umbilical cord blood-derived very small embryonic/epiblast-like stem cells (VSELs),” *European journal of Haematology*, vol. 84, no. 1, pp. 34–46, 2010.

Appendix A

Background Material

A.1 Cell Preparation in Chapter 5

SH-SY5Y cells were from the American Type Culture Collection (ATCC, Manassas, Va, USA) and cultured using a 1:1 mixture of Eagle's Minimum Essential Medium supplemented with F12 Medium containing 10% v/v fetal bovine serum and penicillin/streptomycin mixture. Cultures were seeded at 20% cell density and then allowed to grow for approximately 48 h until they had reached 70% confluence of adherent cells. At this point, staurosporine was added to a final concentration of 5 μM to make the treated SH-SY5Y cells group, while an equivalent volume of PBS was added to the control SH-SY5Y cells group without staurosporine. Both groups were allowed to incubate for 48 h, and then cells were fixed with 10% p-formaldehyde for 15 minutes. Cells were sampled for each of the two conditions and were counted using a hemocytometer to determine cell density present for each aliquot. Undiluted samples had a concentration of approximately 10^6 to 10^7 cells/ml and were diluted to a final assay concentration of around 3000 cells/ml so that light scattering patterns of single cells could be obtained.