TALLINN UNIVERSITY OF TECHNOLOGY
DOCTORAL THESIS
30/2018

Classification and Denoising of Objects in TEM and CT Images Using Deep Neural Networks

ANINDYA GUPTA

TTÜ PRESS

TALLINN UNIVERSITY OF TECHNOLOGY
School of Information Technologies
Thomas Johann Seebeck Department of Electronics
This dissertation was accepted for the defence of the degree of Doctor of Philosophy in Electronics and Telecommunication on May 13, 2018.

Supervisors: Professor Olev Märtens
Thomas Johann Seebeck Department of Electronics
Tallinn University of Technology, Tallinn, Estonia

Professor Yannick Le Moullec
Thomas Johann Seebeck Department of Electronics
Tallinn University of Technology, Tallinn, Estonia

Associate Professor Ida-Maria Sintorn
Center for Image Analysis
Department of Information Technology
Uppsala University, Uppsala, Sweden

Consultant: Dr. Tõnis Saar
Elestron OÜ, Tallinn, Estonia

Opponents: Professor Hannu Eskola
Quantitative Medical Imaging Group
Faculty of Biomedical Sciences and Engineering
Tampere University of Technology, Finland

Associate Professor Lasse Riis Østergaard
Medical Image Analysis Group
Department of Health Science and Technology
Aalborg University, Denmark

Defence of the thesis: June 15, 2018, Tallinn

Declaration:
Hereby I declare that this doctoral thesis, my original investigation and achievement, submitted for the doctoral degree at Tallinn University of Technology has not been submitted for doctoral or equivalent academic degree.
Anindya Gupta

ISSN 2585-6898 (publication)
ISBN 978-9949-83-263-7 (publication)
ISSN 2585-6901 (PDF)
ISBN 978-9949-83-264-4 (PDF)

TALLINNA TEHNIKAÜLIKOOL
DOKTORITÖÖ
30/2018

Objektide klassifitseerimine ja müratustamine TEM ja KT kujutistelt sügavate närvivõrkude abil

ANINDYA GUPTA

TTÜ KIRJASTUS

To my grandfather

Never stop dreaming, never stop believing, never give up,
never stop trying, and never stop learning.

— Roy T. Bennett, The Light in the Heart

TABLE OF CONTENTS

LIST OF PUBLICATIONS

The work of this thesis is based on the following publications; copies of these publications (A-E) can be found in Appendix:

I **Gupta A**, Suveer A, Lindblad J, Dragomir A, Sintorn I-M, and Sladoje N. Convolutional Neural Networks for False Positive Reduction of Automatically Detected Cilia in Low Magnification TEM Images. In *Proceedings of the 20th Scandinavian Conference on Image Analysis (SCIA)*, Tromsø, Norway, June 2017, LNCS-10269, 407-418. doi: /10.1007/978-3-319-59126-1_34

II Bajić B*, Suveer A*, **Gupta A***, Pepic I, Lindblad J, Sladoje N, and Sintorn I-M. Denoising of Short Exposure Transmission Electron Microscopy Images for Ultrastructural Enhancement. In *Proceedings of the 15th International Symposium on Biomedical Imaging (ISBI)*, Washington D.C., USA, April 2018.

III **Gupta A**, Saar T, Märtens O, and Moullec YL. Automatic detection of multi–size pulmonary nodules in CT images: Large-scale validation of a multilayer perceptron based false-positive reduction step. *Medical Physics*, 2018;45(3):1135-49. doi: 10.1002/mp.12746

IV **Gupta A**, Saar T, Märtens O, Moullec YL, and Sintorn I-M. Detection of Pulmonary Micronodules in CT Images and False Positive Reduction Using 3D Convolutional Neural Networks. *submitted for Journal Publication.*

V Lidayová K*, **Gupta A***, Frimmel H, Sintorn I-M, Bengtsson E, Smedby Ö. Classification of Cross-sections for Vascular Skeleton Extraction Using Convolutional Neural Networks. In *Proceedings of the 21th Medical Image Understanding and Analysis (MIUA)*, Edinburgh, Scotland, July 2017, CCIS-723, 182-194. doi: 10.1007/978-3-319-60964-5_16

All papers published or accepted for publications are reproduced with permission from the publishers.

**These authors have contributed equally.*

RELATED PUBLICATIONS

In addition to the papers included in this thesis, the author has also written or contributed to the following publications:

Peer-reviewed

i **Gupta A**, Saar T, Märtens O, and Le Moullec Y. Unsupervised Feature Mapping via Stacked Sparse Autoencoder for Automated Detection of Large Pulmonary nodules in CT Images. *Elektronika ir Elektrotechnika*, 2017; 23(6):59-63.

ii **Gupta A**, Märtens O, Le Moullec Y, and Saar T. A Tool for Lung Nodules Analysis based on Segmentation and Morphological Operation. In *Proceedings of 9^{th} International Symposium on Intelligent Signal Processing (WISP)*, May 2015, IEEE, 1-5.

iii **Gupta A**, Märtens O, Le Moullec Y, and Saar T. Methods for Increased Sensitivity and Scope in Automatic Segmentation and Detection of Lung Nodules in CT Images. In *Proceedings of the 21^{th} International Symposium on Signal Processing and Information Technology (ISSPIT)*, UAE, December 2015, IEEE, 375-380.

Others

iv **Gupta A**, Märtens O, and Le Moullec Y. A Preliminary Computer-Aided Technique for CT Based Lung Segmentation. In *Annual Conference of the Estonian Doctoral School of ICT*, Tallinn, Estonia, December 2014

v **Gupta A**, Märtens O, and Le Moullec Y. A Survey on Open-Access Reference Image Databases for Lung Cancer. In *Annual Conference of the Estonian Doctoral School of ICT*, Tallinn, Estonia, December 2014.

vi **Gupta A.**, Suveer A., Lindblad J., Dragomir A., Sintorn I.-M., and Sladoje N. False Positive Reduction of Cilia Detected in Low Resolution TEM Images Using a Convolutional Neural Network. In *Proceedings of the Swedish Symposium on Image Analysis (SSBA)* , Linköping, Sweden, March 2017.

AUTHOR'S CONTRIBUTIONS TO THE PUBLICATIONS

For Paper I, Gupta A is the main contributor to the design, implementation, training, and testing of the convolutional neural networks. Gupta A conducted the experiments and the evaluation. Gupta A did the evaluation and comparison of the proposed method with the previous method discussed in the paper. Gupta A wrote the paper with input and feedback from the co-authors.

For Paper II, Gupta A contributed to the design and development of the methodology. Gupta A is the main contributor to the convolutional neural networks design, training and testing. Gupta A participated in the evaluation and comparison of the proposed method with three state-of-the-art methods. Gupta A contributed to writing the paper with input and feedback from the co-authors.

For Paper III, Gupta A is the main contributor to the conceptual design and implementation of multiple sizes of pulmonary nodules detection and false positive reduction part of the paper. Gupta A conducted the experiments and the evaluation on four publicly available datasets. Gupta A did the comparison of the proposed method with the state-of-the-art methods. Gupta A wrote the paper with input and feedback from the co-authors.

For Paper IV, Gupta A conceptualized the idea with the help of his co-supervisor. Gupta A is the main contributor to the design and development of the automated detection framework. Gupta A implemented, trained, and tested the 3D convolutional neural networks. Gupta A wrote the paper with input and feedback from the co-authors.

For Paper V, Gupta A implemented the convolutional neural networks and performed the training and testing. Gupta A contributed to the evaluation and comparison of the proposed method with the previous methods discussed in the paper. Gupta A contributed to writing the paper with input and feedback from the co-authors.

ABBREVIATIONS

ANN	Artificial Neural Networks.
AUC$_{PR}$	Area Under the Precision-Recall Curve.
AUC$_{R}$	Area Under the ROC Curve.
BGD	Batch Gradient Descent.
BN	Batch Normalization
CAD	Computer-Aided Detection.
CNN	Convolutional Neural Networks.
CPM	Competition Performance Metric.
CT	Computed Tomography
CTA	CT Angiography
DL	Deep Learning.
DNN	Deep Neural Networks.
ERS	European Respiratory Society.
ESR	European Society of Radiology.
FDA	Food and Drug Administration.
FN	False Negative.
FOV	Field of View.
FP	False Positive.
FPR	False Positive Rate.
FROC	Free-Response Operating Characteristic.
GD	Gradient Descent.
GLCM	Gray Level Co-occurrence Matrix.
GPU	Graphical Processing Units.
HM	High-Magnification.
HU	Hounsfield Units.
ICS	Immotile Cilia Syndrome.
LDCT	Low Dose Multi-Slice Computed Tomography.
LIDC/IDRI	Lung Image Database Consortium/Image Database Resource Initiative.
LM	Low-Magnification.
MAE	Mean Absolute Error.
MBGD	Mini-Batch Gradient Descent.

MLP	Multi-layer Perceptrons.
MM	Mid-Magnification.
MSE	Mean Square Error.
MSLE	Mean Squared Logarithmic Error.
NCC	Normalized Cross-Correlation.
NELSON	Dutch-Belgian Screening Trial.
NLST	National Lung Cancer Screening Trial.
NN	Neural Networks.
PCD	Primary Ciliary Dyskinesia.
PSNR	Peak-Signal-to-Noise Ratio.
ReLU	Rectified Linear Unit.
ROC	Receiver Operating Characteristic
ROI	Region of Interest.
SGD	Stochastic Gradient Descent.
SSE	Sum of Square Error.
SSIM	Structural Similarity Index Measure.
SVM	Support Vector Machine.
TEM	Transmission Electron Microscopy
TM	Template Matching.
TN	True Negative.
TP	True Positive.

LIST OF FIGURES

LIST OF TABLES

INTRODUCTION

The advent of digital images has facilitated the use of computer systems for improved characterization of underlying biological or anatomical structures. Until the 1990s, biomedical and medical images were typically analyzed using low-level image processing methods and mathematical models to solve particular radiological or clinical tasks [1–3]. Although these often semi-automated methods enabled radiologists/clinicians to analyze different abnormalities, they were limited due to computational power and the requirement of constant manual intervention. Lately, increased computational power and evolution of: *image acquisition*, *image analysis*, and *neural networks-based methods* has revolutionized the landscape of automated analysis frameworks. The technological advancements of imaging devices have also elucidated the biological and anatomical behavior of previously unknown complex processes. However, this successful digital transition of imaging devices has also uncovered numerous challenges for the clinicians as well as the image analysts and researchers.

An example of such digital imaging devices is transmission electron microscopy (TEM) which allows structural analysis of biological samples at the nm scale. The comprehensive structural analysis is crucial to extract clinically relevant information. Such analysis typically follows a manual diagnostic procedure, which is labor-intensive, monotonous, error-prone, and time-consuming. For instance, to diagnose a rare genetic disorder: *Primary Ciliary Dyskinesia (PCD)*, pathologists commonly analyze around 50 perfectly perpendicularly cut cilia structures in several high-resolution (~ 1 nm) TEM images. Manually acquiring such images while navigating through a huge search space to cover the whole sample is impractical, requiring two hours/patient. On the other hand, it is viable to steer and detect plausible cilia regions at low-magnification (LM), followed by the acquisition of high-magnification (HM) images only of the detected regions. However, manual detection of cilia at LM (an inevitable requisite for automation of PCD analysis) is itself a challenging task due to inadequate ultrastructural information and high similarities to the non-cilia (Publication A). Another possibility to digitally improve and automate the imaging and analysis process regarding resolution, speed and risk (to damage/destroy the sample) is to acquire, denoise and register short exposure images. This will minimize the influence of imaging artifacts on the automated or the manual diagnostic procedure (Publication B).

Another imaging modality, computed tomography (CT), provides a high-contrast resolution and volumetric characterization of anatomical structures as small as 1 mm. The multi-planer reformation in CT benefits radiologists to simultaneously interpret the anatomical structures in three orthogonal planes, i.e., axial, coronal, and sagittal. In a usual clinical setting, radiologists often analyze 100-500 cross-sectional images of a single volume to conclude some decision about abnormalities. Interpreting such an amount of images, e.g., thoracic CT scans during large-scale cancer screening trials as well as in routine practices to detect early-stage *pulmonary nodules*, is laborious, monotonous, and could take up to 10-15 minutes/scan (Publication C). The diagnostic procedure becomes even more challenging due to the high variability among pulmonary nodules, and their high similarity to the blood vessels when visually analyzed in a 2D slice by slice fashion. Another challenging task for radiologists is to quantify and detect the *vascular pathologies* in CT angiography (CTA) images. CTA imaging is often used for volumetric characterization of blood vessels in the whole body by injecting a contrast medium intravenously, resulting in a large amount of complex data (Publication E).

Given the enormous amount of data and complexity from, e.g., the aforementioned imaging techniques, development of the automated analysis framework is indeed of high desire to assist clinicians in the otherwise error-prone and time-consuming processes.

1. COMPUTER-AIDED DETECTION (CAD)

Computer-aided detection (CAD) is an umbrella term that covers a broad spectrum of the research area in clinical, biomedical and medical imaging. The foremost objective of a CAD framework is to assist clinicians in detecting abnormalities. Clinicians are required to analyze enormous amounts of data in relatively short time, which might lead to erroneous outcomes. Studies have shown that clinicians occasionally misinterpret some visible abnormalities [4–9]. Such errors may lead to a perilous repercussion on patient's health. Therefore, assimilation of CAD frameworks is worthwhile to reduce the interpretation time and also the detection errors.

The CAD framework can be perceived as a pattern recognition system that impersonates the human observers to perform any specific task [10]. Pattern recognition is a technical perspective associated with the scientific field of *Artificial Intelligence*. Pattern recognition is dedicated to developing programs that enable computers to learn the underlying patterns, and thereby to make reasonable decisions using those learned patterns. For example, to identify impairments in thoracic CT scans, the CAD will quest for the size and shape patterns of nodules; or to perform a PCD analysis, the CAD will quest for the textures and ultrastructural patterns of potential cilia. Indeed, strategized modular components are essential requirements for a CAD to comprehend such complex learning. In a sequential setting, a traditional CAD system can typically be composed of three modules: *image preprocessing*, *candidate-screening*, and *classification* [2].

The *image preprocessing module* usually consists of the procedures to enhance the image quality. This module aims to reduce the influence of artifacts caused by the imaging devices. Typical preprocessing steps are noise removal, normalization of intensity non-uniformities, isotropic interpolation, etc.

The *candidate-screening module* includes techniques to segment and locate structures of interest in the preprocessed image. Segmentation is commonly used to separate objects from the background or to detect pertinent structures in the image. This module aims to locate a substantial amount of plausible structures while rapidly screening through the entire image.

The *classification module* aims at classifying the structures using a set of discriminative features within an empirically optimized classifier. The features are often task-oriented and are computed from the characteristics of the segmented regions. For instance, circularity, elongation, contrast, homogeneity,

and spiculation[1] are some typical features of interest for classifying cancerous nodules and non-nodules. However, not every feature possibly exhibits the same discriminative power or could even be pernicious when performing the same task for solving two different application challenges. For example, circularity is a useful shape feature to discriminate between nodules and non-nodules; however, it will not have the same usefulness while discriminating cilia from non-cilia since both structures tend to be circular. Hence, careful selection of extracted features is a crucial step prior to classification. Once the discriminative features are selected, they are fed as an input to the classifier. In an iterative learning process (training) using a substantial amount of labeled data, the classifier determines the optimal boundaries for classifying structures/classes (e.g., normal and abnormal) in the multi-dimensional feature space. Since it involves learning from the labeled data, this learning procedure is typically referred to as supervised learning. During testing, the classifier predicts the unlabeled input structures as one of the classes.

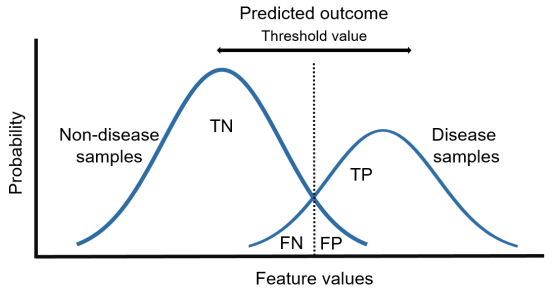


Figure 1.1 The distribution of the predictions obtained by a classifier. For every possible threshold value, the classifier discriminates the samples into true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN) fractions.

Once the output from the classifier is obtained, the significance of the CAD system is determined using different evaluation metrics. The predictions obtained by the classifier are typically divided into four fractions and follow a distribution as shown in Fig. 1.1. The true positives (TP) and true negatives (TN) fractions correspond to the samples correctly classified as abnormal and normal, respectively. The false positives (FP) and false negatives (FN) fractions correspond to the samples wrongly classified as abnormal and normal, respectively. Using these fractions, the performance of a classifier to discriminate abnormal and normal cases is often determined by the following statistics:

$$Sensitivity, Recall = \frac{TP}{TP + FN} \,, \tag{1.1}$$

$$Specificity = \frac{TN}{TN + FP} \,, \tag{1.2}$$

[1]spiculation is associated to the subjective assessment of the malignancy likelihood

$$Precision = \frac{TP}{TP + FP} \ , \tag{1.3}$$

$$F\text{-}score = 2 \times \frac{Precision \ \times \ Recall}{Precision + Recall} \ , \tag{1.4}$$

In this thesis, the receiver operating characteristic (ROC) analysis and the free-response operating characteristic (FROC) [11] analysis are also used to evaluate the performance of a CAD framework. The ROC curve plots the sensitivity as a function of the false positive rate (1-specificity) for all possible threshold points. The area under the ROC curve (AUC_R) is a typical evaluation metric derived to determine the performance of a classifier. An AUC_R score of one corresponds to perfect discrimination whereas the score of 0.5 corresponds to random guessing. The FROC [12] curve is not limited to an upper bound on the negative object axis, i.e., false positive rate (FPR). It plots the sensitivity as a function of the average number of FPs per image and has higher statistical discriminative power [12]. It is more sensitive at detecting small differences between performances when multiple abnormal regions are present in a single image.

Considering the profundity of the detection associated tasks, it is crucial that the CAD systems manifest a high sensitivity with as low FPR as possible. The FPs are responsible for the potential detection errors, resulting in increased analysis time, while the FNs represent missed detections, which may have fatal impact. Although each module has its importance, the classification module is predominantly the most pivotal in this sequential setting when determining the overall significance of the CAD. Generally, the methods employed in the candidate-screening module yield a high sensitivity but at a high FPR. Conversely, the classification module attempts to reduce the high FPR while maintaining the high sensitivity, and thereby potentially called-for rigorous learning-based methods to overcome the complexities induced by the former module.

1.1. Deep Neural Networks for CAD

Due to the involvement of machine learning methods such as neural networks, random forests, support vector machine (SVM), etc. for learning the discriminative features in the classification module, this approach is often referred to as feature-based machine learning [13]. Until recently, feature-based learning remained as the critical approach for the classification. Acknowledging that the extraction and selection of task-oriented features at hand are intensively exhaustive and challenging, CAD practitioners started to incline toward generic methods that rely on learning from the raw data. Such image-based machine learning approaches (or end-to-end learning approaches) [13] directly use the raw pixel values of an image as features for the classification instead of explicit computation of task-oriented features. Although this perspective of exploiting

data benefited the practitioners profoundly, it was still confined due to low computational power and training data [2, 14].

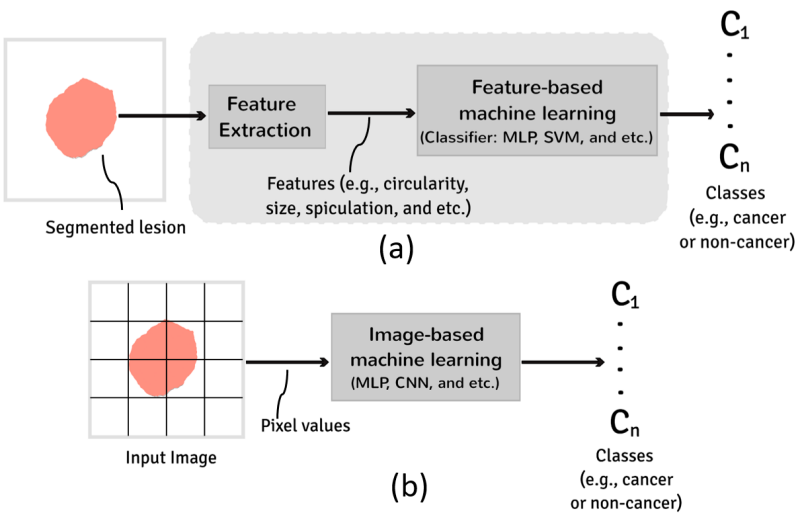


Figure 1.2 CAD systems for classifying lesions. (a) Feature-based, and (b) Image-based. The prime distinction between them is direct (or end-to-end) learning from pixel values.

Lately, increasing computation power and evolution of neural network-based learning approaches have exalted the CAD paradigm. The deep neural networks (DNN)-based learning approaches, and specifically, convolutional neural networks (CNN) are commonly referred to as deep learning methods. Although the existence of DNN dates back to the 1970's [15] and they were already exploited in 1995 [16] for medical image analysis, they were not widely recognized until 2012 [17]. Since then, they have become the foremost technique of interest for many computer vision and image analysis problems. The impulse of employing CNN is indeed galvanized by many factors such as data augmentation, development of new techniques for the training of sophisticated DNN, and parallel computing using graphical processing units (GPU). The schematic of the CAD systems using traditional feature-based machine learning and image-based machine learning (or DNN) are illustrated in Fig. 1.2.

The inception of deep learning methods has rationalized the perspective of the CAD systems. CAD researchers have conspicuously experienced many advantages by leveraging the DNN-based methods over feature-based learning methods. Apart from the automated learning of discriminative features, they can also simultaneously locate multiple impairments in a particular image. Additionally, DNN-based methods have also preluded the concept of fine-tuning, allowing the inference of the pre-trained models for the domain-agnostic problems. For instance, the *VGG16* – a popular DNN-based method, trained on the ImageNet (one of the largest annotated dataset of natural images) to

classify 1000 classes can be fine-tuned to comprehend the underlying features of pulmonary nodules in CT images for binary classification [18].

1.2. Problem Statement and Research Objectives

Aiming at addressing the aforementioned challenges, this thesis focuses on the development of automated analysis frameworks for digital images captured using the TEM and the CT imaging modalities. This has been accomplished both by pushing the limits of well-established techniques such as feature-based methods, as well as by penetrating the contemporary pattern recognition techniques such as DNN-based methods.

The primary goal of this thesis is to develop and compare computer-aided detection (CAD) framework solutions ranging from traditional image analysis and machine learning methods to modern DNNs-based representation learning methods. This involves implementation and applications of DNNs in the classification of objects in medical (CT) images, as well as in the classification and denoising of objects in biomedical (TEM) images.

To this end, four radiological or clinical problems related either to image analysis or acquisition are investigated by setting up the following objectives:

1 To develop a DNN-based method for classifying automatically detected cilia in low magnification TEM images;

2 To develop a DNN-based method for ultrastructural enhancement by denoising short exposure TEM images;

3 To develop a traditional image analysis-based method for the detection of different sizes of pulmonary nodules in CT images and to develop a DNN-based method for classifying them;

4 To develop a traditional image analysis-based method for the detection of pulmonary micronodules in CT images and to develop a DNN-based method for classifying them;

5 To develop a DNN-based method for classifying cross-sections of vascular skeletons in CTA images.

1.3. Contribution of This Thesis

With deluge of data in the current clinical practices, the clinicians are likely to misinterpret subtle abnormalities, resulting in erroneous diagnostic observations. Thus, it is vital to facilitate the clinicians with resilient CAD systems in the time-consuming, labor-intensive and error-prone tasks. However, existing CAD systems are restrained due to the heterogeneity induced by the biological or

anatomical structures and low computation power. Until recently, designing highly discriminative feature sets was considered as one of the involved challenges in the CAD research.

Lately, the renaissance of DNN-based methods as a compelling technique has unequivocally remolded the CAD systems. As stated earlier, this thesis aims at rejuvenating the CAD systems using DNN-based methods. Given the general framework mentioned in Section 1, this thesis mainly focuses on providing solutions for the classification and denoising modules using DNN-based methods. Acknowledging the potency of the deep neural networks, potential solutions have been contemplated for the challenges imposed by both TEM and CT imaging modalities.

The main contributions of this PhD thesis are summarized as follows:

1 A CNN classifier is developed to reduce the false positives detected by an existing template matching (TM) method in low-magnification TEM images. Given the small amount of training data, curriculum learning and data augmentation techniques are applied to improve the performance of the classifier. The framework is tested on multiple sets of images. Adding a CNN classifier improved the overall F-score from 0.47 to 0.59.

2 Aiming at the restoration of the short exposure HM MiniTEM™[2] images, a novel multi-stream CNN module is developed and compared with three state-of-the-art denoising methods. Techniques such as batch normalization and residual learning are harnessed to improve the overall performance of the CNN module. Using a large set of 100 image sequences, three experimental studies are conducted to determine the optimal denoising strategy. The proposed CNN module is only trained for the first experimental study and used as it is for the other two studies to manifest the transfer learning aspect of deep learning (DL). The presented CNN model achieved an improved peak-signal-to-noise ratio (PSNR) of 40.84 dB.

3 Given the heterogeneity among pulmonary nodules in CT scans, an automated CAD system is developed for the early manifestation of lung cancer. Methods for both lung segmentation and nodule detection in a candidate-screening module are developed using traditional image analysis methods. An upgraded voxel-based feature extraction approach is developed to discriminate the candidates using a multi-layer perceptrons (MLP) classifier. The proposed CAD system is evaluated on altogether 1052 CT scans taken from four publicly available datasets. The presented CAD system achieved an overall sensitivity of 85.6% with only 8 FPs/scan.

[2] *A tabletop low-voltage TEM hardware solution from* Vironova AB, Stockholm, Sweden

4 An automated CAD system is developed for the detection of micronodules in CT scans. Methods for both lung segmentation and nodule detection in a candidate-screening module are developed using traditional image analysis methods. A novel 3D CNN model for automatic feature extraction is developed and compared with traditional hand-crafted feature extraction techniques. The methods are evaluated on altogether 598 CT scans taken from the largest publicly available dataset, and achieved an overall sensitivity of 86.7% with only 8 FPs/scan.

5 The existing knowledge-based filters for the vascular skeleton extraction generate a large number of false positive nodes. Aiming at simplified extraction workflow, a patch-based CNN classifier is developed to classify the cross-sections of multi-size arteries. Using 25 CTA volumes of the lower limbs, the performance of the developed CNN classifier is evaluated and compared to the existing method. The workflow employing a CNN classifier generates the final vascular skeleton in a single algorithm pass, thereby eliminating the requisite to locate the skeletons of small arteries in the subsequent iteration. Adding a CNN classifier improved the overall F-score from 0.43 to 0.82.

1.4. Thesis Outline

This Ph.D. thesis is divided into 5 chapters. After introducing the pathological and radiological challenges associated with the TEM and the CT imaging modalities, this chapter elucidates the concept of CAD systems and its generic modules. Next, it also covers the transition of the classification module from the feature-based machine learning to the image-based machine learning methods. Given the challenges from the conventional feature-based machine learning methods, this chapter also briefly explains how DNN-based methods have benefited the CAD researchers.

Chapter 2 presents the technical perspective of DL. It explains underlying concepts of the DNNs used in this thesis and their existing variants in the current practices. It also highlights the commonly known loss functions, optimization algorithms, and regularization techniques for DNN used in this thesis.

Chapter 3 focuses on two problems associated with current manual TEM imaging. Firstly, FP reduction of cilia detected in low magnification TEM images using CNN. At LM level, the non-cilia candidates exhibit high similarities to cilia candidates. It is hence vital to employ learning-based methods such as CNN for fast and effective automated image analysis. Secondly, denoising of short exposure high-magnification TEM images for ultrastructural enhancement using a CNN. Acquiring short exposure images is required to minimize the issues of imaging artifacts, however, at the cost of signal-dependent noise.

Denoising is a well-studied ill-posed problem. Aiming at preserving the structural information, three state-of-the-art methods are exploited and compared with a novel multi-stream denoising CNN explicitly developed for the short exposure images acquired using the low-voltage table-top MiniTEM™.

Chapter 4 tackles the problem of automatic detection of multi-sizes pulmonary nodules in CT scans. To do so, two CAD systems are developed. The first CAD system presents a conventional feature-based CAD system for the early manifestation of lung cancer. Until now, the proposed CAD system is the only CAD system that has been tested on four publicly available datasets. The second CAD system employs a 3D CNN to detect micronodules for the manifestation of silicosis. Manual interpretation of micronodules in CT images is labor-intensive, erroneous, and time-consuming. Given the success of CNN, it is worthwhile to exploit the 3D CNN for detection of the micronodules. This is the first study to be reported on the automatic detection of micronodules in the Lung Image Database Consortium/Image Database Resource Initiative (LIDC/IDRI) dataset.

Chapter 5 focuses on a CNN-based solution to the problem associated with the extraction of vascular skeletons in the CT angiography images. Current practices of delineating vascular skeletons seek for fast, automated, and simplified extraction techniques. Conventional image analysis methods suffer from a large amount of multi-size FP; thereby CAD researchers potentially seek for rigorous learning-based methods such as CNN.

Finally, the thesis is concluded by highlighting the possible future perspectives. Figure 1.3 shows an infographic overview of this Ph.D. thesis.

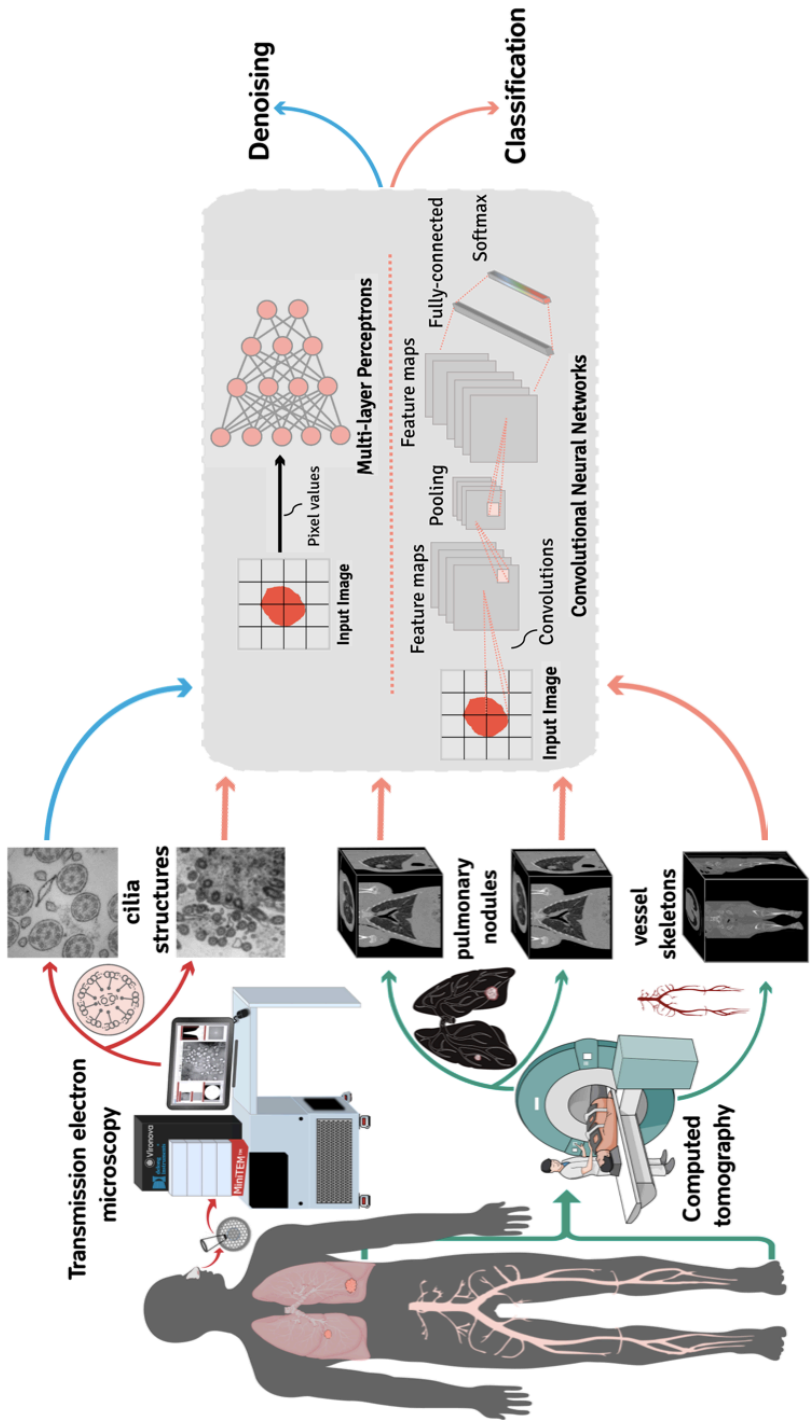


Figure 1.3 Infographical overview of the overall PhD thesis work

2. BACKGROUND: DEEP NEURAL NETWORKS

In 1943, neurophysiologist Warren McCulloch and mathematician Walter Pitts provided an insight into the functionality of the brain by mimicking its working mechanism, and consequently formalized the first computational model of neural networks (NN) [19]. Their mathematical model transpired a new scientific perspective- artificial neural networks (ANN).

Inspired by that Bernard Widrow and Marcian Hoff conceptualized the first ANN —*ADALINE* for a real-world problem in 1959 [20]. However, the research slowed down in the 1970's since a perceptron was not capable enough to approximate functions outside a very narrow class [21]. By 1986, some of the limitations were overcome and the interest in NN rejuvenated because of the backpropagation algorithm[1] [22]. Henceforth, ANN are continually proving themselves as a very effective and a powerful tool to solve complex tasks. Figure 2.1 shows the milestones in the evolution of ANN.

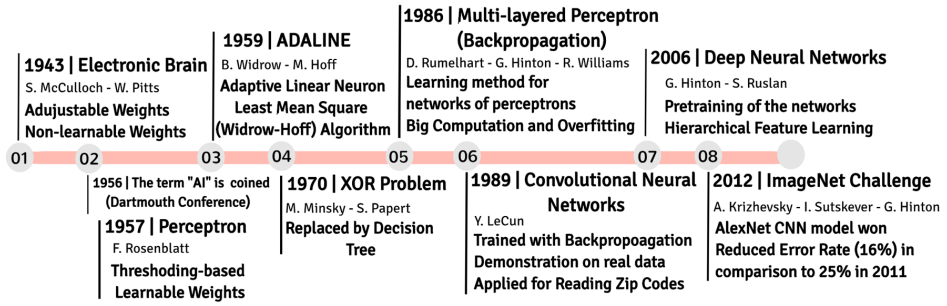


Figure 2.1 Milestones in the development of deep neural networks (DNN)

2.1. Artificial neural networks

An ANN model is inspired by the biological nervous systems, which consists of a group of artificial neurons (or perceptrons). A perceptron is a single processing entity comprised of some functions such as partial summation, a *bias*

[1]Lesser known fact about backpropagation: The minimisation of errors through gradient descent is even dated back to the 1847. `http://people.idsia.ch/~juergen/who-invented-backpropagation.html`

to control the influence of perceptrons, and an *activation function* (explained in Section 2.4) to stimulate a nonlinear behavior. Several such weighted neurons are interconnected with each other. The *weight* parameters define the connection strength among these neurons, as shown in Fig. 2.2.

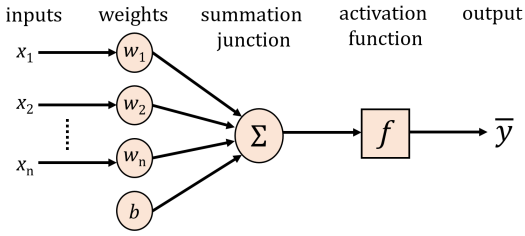


Figure 2.2 A single neuron with inputs $x_{i=1,\cdots,n}$, bias b, non-linear activation function $f(\cdot)$, and predicted output $\bar{y}$.

In this canonical process, each input data point x_i is multiplied by its corresponding weight parameter w_i, followed by the summation of weighted inputs and bias b. The bias parameter adds an offset to the data. This linear combination is transformed by a non-linear activation function $f(\cdot)$ to predict the output. For instance, the neuron with multiple inputs $x_{i=1,\cdots,n}$ computes the output $\bar{y}$ as follows:

$$\bar{y} = f\Big(\sum_{i=1}^{n} w_i x_i + b\Big)\,, \tag{2.1}$$

where the parameters $w_{i=1,\cdots,n}$ are weights, b is *bias* and $f(\cdot)$ is a non-linear *activation function*, also refered to as a transfer function.

2.2. From Perceptrons to Multilayer Perceptrons

A perceptron is a simple algorithm that can only solve linearly separable problems [21]. However, cascading of several such neurons in multiple layers forms a richer hierarchical model commonly known as multilayer perceptrons (MLP) where all neurons in the previous layer are densely (or fully)-connected[2] to the neurons of succeeding layer. For example, the MLP in Fig. 2.3(a) consists of three densely-connected layers, i.e., an input layer, a hidden layer, and an output layer. For a given set of inputs $x_{i=1,\cdots,n}$ with one hidden layer $h_{j=1,\cdots,m}$, the output $\bar{y}$ can be computed as follows:

$$\bar{y} = f_{out}\Big(\sum_{j=1}^{m} w_j h_j + b_1\Big), \quad h_j = f_h\Big(\sum_{i=1}^{n} w_i x_i + b_0\Big)\,, \tag{2.2}$$

where parameters w_i, b_0 and $f_h(\cdot)$ are respectively the weights matrices, the *bias*, and the transfer function associated to input and hidden layers; w_j, b_1, and

[2]The connectivity pattern of a DNNs is referred to as the network's architecture.

$f_{out}(\cdot)$ are the weights matrices, the *bias*, and the transfer function associated to hidden and output layers. The neural networks model with a depth of more than two layers are generally referred to as a *Deep neural networks* (DNN).

2.3. Learning Process

Before introducing the currently most popular variation of DNN, i.e., convolutional neural networks (CNN), it is worthwhile to get acquainted with the underlying learning configuration that makes them so powerful. Both MLP and CNN follow the same learning process.

The DNN aim to solve a particular task through learning from a given set of d instances, $d \in \{1, 2, \cdots, p\}$, which consists of an input vector $x_d = [x_1, x_2, \cdots, x_n]$ and corresponding label y_d. The process of learning is typically referred to as training process and often associated with the supervised learning. This thesis is based on supervised learning approaches. In a supervised learning scenario, when discrete outputs are desired (as in Papers I, III-V), the task is associated to a classification problem whereas when continuous outputs are desired (as in Paper II), the task becomes more of a regression problem. The training process, as shown in Fig. 2.3, includes multiple steps: *input normalization*, *weights initialization*, *forward propagation*, *loss function evaluation*, *backward propagation*, and *weights update*.

The *input normalization* is a transformation performed as a data preprocessing step to standardize the range of data points in a close bounded range. This step is essential to prevent the network from generalizing towards dominating features. In this thesis, the normalization is performed by subtracting the minimum data point variable and dividing by the difference between maximum and minimum data point variables, resulting in normalized values in the range [0,1]. In addition, the z-score standardization is also performed in this thesis by subtracting the mean and dividing by the standard deviation, resulting in a standard normally distributed data.

The *weights initialization* step assigns random weights to neurons as a starting point. The initial values of weights significantly influence the learning process. Large weight values can saturate the transfer function, causing complete loss of gradient[3] through saturated neurons (or exploding the gradients). Small weight values can result in very small gradients, causing the vanishing gradient problem (discussed in Section 2.4). The weights should not be assigned symmetrically to not receive the same updates during training. As suggested in [23], the weights w should be initialized using a Gaussian distribution with zero mean and variance:

$$variance(w) = \frac{2}{n_{in} + n_{out}}\ , \tag{2.3}$$

[3]Partial derivatives are often referred to as gradients by the deep learning community. It is a measure of how much the error changes with respect to a change in a weight or bias value. The gradient at any point is the product of all the previous gradients up to that point when traversing the network backward.

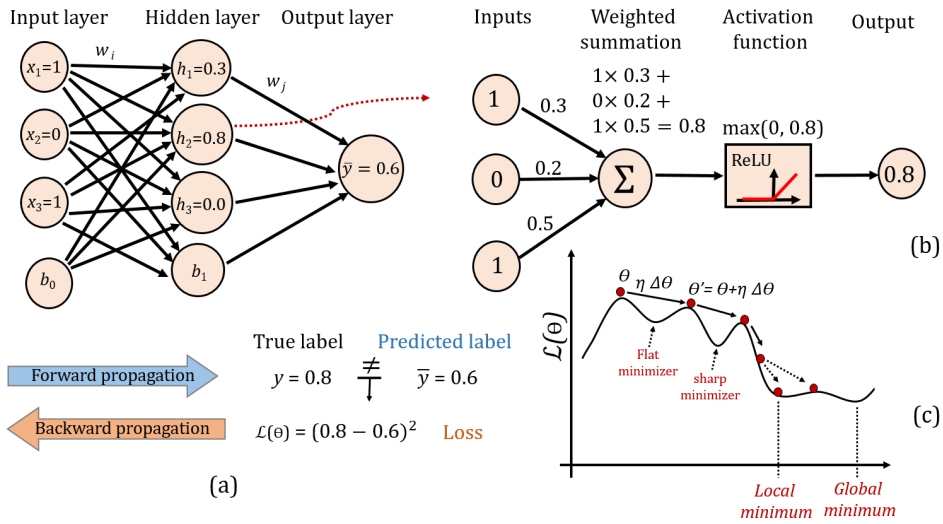


Figure 2.3 Building blocks and learning process of a DNN. (a) An MLP with an input layer, a hidden layer, and an output layer. The neurons in each layer are densely-connected to all neurons of the subsequent layer. Given a set of inputs $x_{i=1,..,3}$ with randomly initialized weights w_i, neuron activations are calculated and propagated forward to the next layer to obtain the output $\bar{y}$. (b) An example is showing the forward propagation by zooming in on one perceptron $h_{j=2}$, which computes the weighted non-negative activation using the ReLU transfer function. (c) Optimization of the loss function $\mathcal{L}(\theta)$ using a gradient-based learning algorithm. In each step, the current weights (red dot) are moved along the steepest direction $\Delta\theta$ (direction arrow) by learning rate (step size) η, where $\theta = (w, b)$. A high learning rate can overshoot and miss an optimum along a steep direction as shown by the dotted arrows. Decaying the learning rate over time allows to explore different domains of the loss function by jumping over valleys at the beginning, and fine-tune parameters with smaller learning rates in later stages.

where n_{in} and n_{out} are respectively the number of inputs and outputs of a corresponding layer. This initialization benefits the training in practice since the weights are sufficiently large to propagate gradients smoothly across the network. However, the choice of weight initialization strategy is rather empirical and varies according to the problem.

The *forward propagation* step propagates the inputs through the hidden layer(s) to calculate the output. In a sequential flow, the input layer propagates non-linear outputs (or activations) to the hidden layer. Using the outputs from the hidden layer as inputs, this process is repeated for the output layer neurons. Figure 2.3(b) shows an example of forward propagation step where the output of a hidden layer neuron is computed using the weighted activations of the input layer.

A *loss function* is essential to evaluate the learning performance of the network. The loss function quantifies the difference between the predicted label

$\bar{y}$ and true label y. After the forward propagation step, the total loss (or error) of the network is estimated using a loss function $\mathcal{L}$. For example in Fig. 2.3(a), the sum of squared error (SSE) loss is computed as:

$$\mathcal{L} = \sum_{d=1}^{p} \frac{1}{2}\Big(y_d - \bar{y}_d\Big)^2 , \tag{2.4}$$

The *backward propagation* step propagates the loss from the output to input layers to calculate the gradients of the loss function. Once the total loss is computed, the gradients are calculated by applying the chain rule for derivatives. For instance, the partial derivatives of the total error with respect to the weight w_i in Fig. 2.3(a) are computed as:

$$\frac{\partial \mathcal{L}}{\partial w_i} = \frac{\partial \mathcal{L}}{\partial \bar{y}} \frac{\partial \bar{y}}{\partial h_j} \frac{\partial h_j}{\partial w_i} , \tag{2.5}$$

where $\dfrac{\partial \mathcal{L}}{\partial \bar{y}}$ corresponds to the partial derivative of total loss with respect to the output of the network, $\dfrac{\partial \bar{y}}{\partial h_j}$ corresponds to the partial derivative of the output with respect to the j^{th} neuron in the hidden layer, $\dfrac{\partial h_j}{\partial w_i}$ is the partial derivative of the j^{th} neuron with respect to the i^{th} weight.

The *weights update* step is performed after computing all the gradients of the network during backward propagation. The weights are updated iteratively in the opposite direction of the gradient (w.r.t. some learning rate $\boldsymbol{\eta}$) to find a local (or global) minimum using an optimization (or learning) algorithm. The iterative methods belong to the gradient-based optimization. For instance, Gradient descent (GD) is a popular optimizer, which minimizes the loss by updating the weights so that the difference between a true label and a predicted label is minimized. Learning is performed by taking small steps η in the direction of the slope created by the loss function (Figure 2.3(c)). A high learning rate corresponds to bigger steps and may speed up the learning to converge to an optimal set of weights. However, it could also overshoot and miss an optimal minimum along a steep direction.

2.4. Activation Functions

An activation (or transfer) function $f(\cdot)$ is used to perform a non-linear transformation on the linear perceptrons (or neurons). The choice of activation functions in DNN has a significant effect on the training process and performance. Typically, DNN performs linear operations (e.g., inner product or convolution) on inputs and their weights, which are then followed by a $f(\cdot)$ operation to perform thresholding on the calculated output. The usage of the activation

function depends on the DNN type and also on the type of layer in which they operate.

The *sigmoid* function is a particular type of logistic function, which is often used as an activation function to obtain the output of the hidden layer neurons. It takes real-valued inputs and squashes them monotonically into a range of 0 to 1, i.e., $f(x) \in (0, 1)$ while centering at the value of 0.5. It implies that the large negative values become 0 and large positive values become 1. Given that it has an exponential in its function, the derivative $f'(x)$ can be calculated as shown in Fig. 2.4(a). During backpropagation, the gradients for the neurons whose output is close to 0 or 1 become nearly zero (or minimal), and thus, almost no signal flows through those saturated neurons to their weights. If the initial weights are too small, most of the neurons will be saturated, and the network will not converge to optimal parameters. This issue is commonly referred to as "vanishing gradients".

On the contrary, weights initialized with large values can cause a large change in the loss, and thus the gradients will also grow exponentially to large values. This issue is commonly referred to as "exploding gradients". Exploding gradients can saturate the activation functions and result in an unstable network that can no longer be updated. Therefore, it is critical to initialize weights of sigmoid neurons carefully or to clip the norm of the gradient at some threshold (known as gradient clipping).

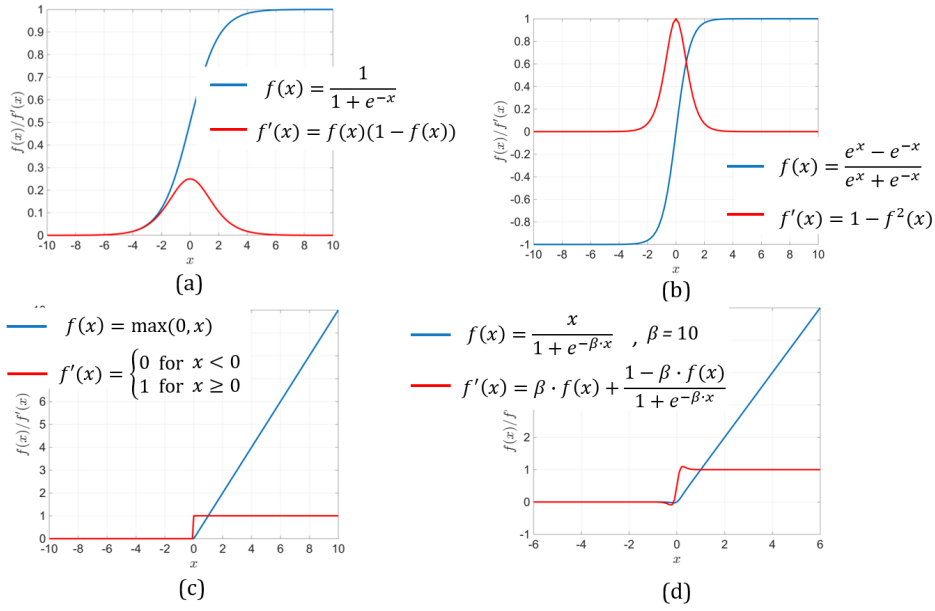


Figure 2.4 Plots of the activation functions with their corresponding derivatives: (a) Sigmoid, (b) Tanh, (c) ReLU, and (d) Swish functions.

The *hyperbolic tangent* (tanh) function is a preferable alternative to the sigmoid function which is also used for hidden layer neuron output. It takes

real-valued inputs and squashes them monotonically into a range of -1 to 1 (Figure 2.4(b)). As for the sigmoid activation, it also suffers from the saturation and vanishing gradient problem.

The *ReLU* function was introduced to address the vanishing gradient problem. It is resistant to this problem at least in the positive region ($x > 0$), which means that the neurons do not propagate all zeros backward to the network. The range of ReLU is between 0 to $+\infty$ (Figure 2.4(c)). Given that it inherits the behavior of a positive linear function, the convergence of SGD is accelerated in comparison to the sigmoid or tanh functions. However, ReLU activated neurons tend to be fragile during training and can be inactive during the entire learning process. For instance, if $x < 0$ during the forward propagation, the neuron remains inactive and thereby kills the gradient while propagating back through the network. Several variations of ReLU are introduced to overcome its limitations such as leaky ReLU [24], and parametric ReLU [25]. Recently, researchers at Google brains introduced a self-gated activation function – *Swish* [26]. This function is a modified version of the sigmoid function and reported to perform better than the variants of ReLU function (Figure 2.4(d)). The ReLU activation function is used in Papers I, II, IV, and V.

The *softmax* function (or classifier) is a generalization of the logistic function. When dealing with classification problems, the linear functions such as ReLU compute unbounded output $\bar{y}$ values. Softmax is a normalized exponential function that squashes the values of each neuron in the output layer to be between 0 and 1 and divides each output in such a way that the total sum of the outputs is equal to 1. The output of the softmax is equivalent to a categorical probability distribution. It is often utilized with negative log-likelihood (or cross-entropy) loss function. The arbitrary values $\bar{y} \in \mathbb{R}^C$ are transformed into normalized probability estimations $p \in \mathbb{R}^C$ to compute softmax for a single instance as:

$$p_k = \frac{\exp \bar{y}_k}{\sum_{i=1}^{C} \exp \bar{y}_i}, \tag{2.6}$$

where $k, i \in \{1, \ldots, C\}$ range over classes, and p_k, $\bar{y}_k$, $\bar{y}_i$ refer to class probabilities and values for a single instance. The Softmax is used in Papers I, III, IV, and V.

2.5. Loss Functions

The loss function has a key role in the optimization of the DNN. The value of the loss function $\mathcal{L}$ shows the discrepancy between the predicted values $\bar{y}$ and true values y. The minimization of the loss implies that a model starts converging to an optimal set of parameters. The loss function is also referred to as empirical risk term and does not contain any trainable parameters. Much like for the activation functions, the choice of the loss function is influenced by the task at

hand. If the task is a linear regression problem, the variants of squared errors can be a suitable loss function. In case of logistic regression such as classification tasks, the cross-entropy error is a more suitable loss function.

Let's consider $\boldsymbol{\theta}$ as the parameters of a model to be learned (or optimized), $f(\cdot)$ represents the activation function and $\mathbf{x}_i = \{x_i^1, x_i^2, \cdots, x_i^m\} \in \mathbb{R}^m$ is a training sample. To introduce generic loss functions, the empirical risk term can be represented as:

$$\mathcal{L}(\boldsymbol{\theta}) = \frac{1}{n} \sum_{i=1}^{n} \left(y_i, f(\mathbf{x}_i, \boldsymbol{\theta}) \right) , \tag{2.7}$$

Mean Squared Error

The mean squared error (MSE) or quadratic loss is often used as a performance measure for linear regression problems. It is computed as:

$$\mathcal{L} = \frac{1}{n} \sum_{i=1}^{n} (y_i - \bar{y}_i)^2 , \tag{2.8}$$

It minimizes the residual sum of squares, i.e., $(y_i - \bar{y}_i)$. However, it suffers from slow convergence when used with sigmoid activation, which is not the case with ReLU or linear activations. This is because when the output of the sigmoid is zero or 1, the derivatives become nearly zero. If the loss $\mathcal{L}$ is too large or too small, the derivatives get closer to zero, and thus, slows down the convergence. The mean squared error is used in Paper II.

Mean Squared Logarithmic Error

The mean squared logarithmic error (MSLE) measures the logarithmic difference of the estimated and true values. It penalizes under-estimated values more than the over-estimated values. It is computed as:

$$\mathcal{L} = \frac{1}{n} \sum_{i=1}^{n} \left(\log(y_i + 1) - \log(\bar{y}_i + 1) \right)^2 , \tag{2.9}$$

Least Squares Error (L_2 - norm)

The L_2 - norm (or regularized expectation loss) minimizes the squared differences between the estimated and existing true values. It is highly sensitive to outliers in the training set because of the squared differences which lead to much larger errors. It is mathematically similar to MSE without a division by n samples. It is computed as:

$$\mathcal{L} = \sum_{i=1}^{n} (y_i - \bar{y}_i)^2 , \tag{2.10}$$

Mean Absolute Error

The mean absolute error (MAE) minimizes the absolute difference between the estimated and existing true values. In comparison to the MSE, it is more robust to outliers since it does not make use of square. It is computed as:

$$\mathcal{L} = \frac{1}{n} \sum_{i=1}^{n} \left| y_i - \bar{y}_i \right| , \tag{2.11}$$

Cross-entropy loss

Cross-entropy loss is useful when dealing with classification problems using DNN. It quantifies the discrepancy between the probability distributions of estimated and true values. In comparison to MSE where sigmoid and softmax activations suffer from saturation and slow learning, the use of cross-entropy loss greatly improves the performance of models with these activations. A large cross-entropy loss means that the difference between two distributions is large whereas small loss implies that two distributions are similar to each other. It is computed as:

$$\mathcal{L} = -\frac{1}{n} \sum_{i=1}^{n} \left[y_i \log(\bar{y}_i) + (1 - y_i) \log(1 - \bar{y}_i) \right] , \tag{2.12}$$

Negative Log Likelihood

Negative log likelihood (softmax loss) is often used to estimate the accuracy of a classifier. It is a soft accuracy measure that incorporates the idea of probabilistic confidence. It is used when the model outputs a probability for each class (binary or multiple classes), rather than just the most likely class. It is computed as:

$$\mathcal{L} = -\frac{1}{n} \sum_{i=1}^{n} \log(\bar{y}_i) , \tag{2.13}$$

The cross-entropy loss function is used in Papers I, III, IV, and V.

2.6. Representation Learning Using Convolutional Neural Networks

Put simply, CNN is a powerful version of MLP. From a biological perspective, the CNN emulates the functionality of the visual cortex, which uses a combination of simple and complex cells to encode richer representations[4] progressively in an image [27]. Similarly, the CNN also employs several convolutional layers (simple cells) using different filters and pooling layers (complex cells) in a hierarchical structure to encode discriminative representations [28]. Since

[4]In the context of convolutional neural networks, the parameters are often synonymized as representations or feature maps.

CNN's are capable of exploiting multi-dimensional data in a matrix form, they are a much-enriched model for classification (or detection) tasks compared to the shallow MLP where the multi-dimensional input is structured into a vector form and, consequently lose the connectivity between local substructures.

The two principal factors of CNN's are parameters sharing and pooling layers. First, CNN's have a benefit in the fact that they are translation-equivariant [5], wherein they share the same filters (also refered to as kernels or parameters) in a local sub-region (or receptive field[6]) of the input layer to encode low-level representations of the objects or region, independently from their positions within an image. Shared parameters drastically reduce parameters that are mapped to the hidden layer, unlike the global receptive field of the MLP where neurons in the hidden layers are densely connected to the input layer. For instance, if an image of $500 \times 500 \times 1$ pixels is given as an input to the MLP, it will have 25 millions parameters ($500 \times 500 \times 1 \times 100$) for 100 neurons in just one hidden layers, and even gets much bigger when multiple layers are cascaded. On the other hand, one convolution layer consisting of 64 feature maps using $5 \times 5 \times 1$ filters will have only 1 664 parameters ($5 \times 5 \times 1 \times 64 + 64$). Second, the subsampling or pooling layers benefit CNN by conferring a certain amount of translation-invariance[5], and spatial-dimensionality reduction, and thus restricting the network from overfitting.

A typical CNN architecture consists of several convolutional layers and pooling layers on top of the dense layers, as shown in Fig. 2.5(a). The convolution layers encode several different representations by convolving over the entire image. The initial convolutional layers comprehend the low-level features such as a circle, an edge, and a vertical line and higher layers encode more complex representations such as textures. These representations are then captured by the activations (or feature maps). Once the representations are extracted, the classification is performed using dense layers.

Convolution Layer

The convolutions are the fundamental operations of the convolutional layer. The convolutional layer consists of several small kernels or filters that are applied to the whole input image to compute the output. To compute the output for a given 2D image $I_{x \times y}$, a set of small kernels k of size $m \times m$ are defined to cover the local receptive fields. The kernels shift over the whole image for computing the output and followed by adding a bias term for each k filter. Finally an activation function $f(\cdot)$ is employed for all of the pixels of the output images to induce nonlinearity. An example is shown in Fig. 2.5(b). The single output image

[5] *One common misapprehension:* Convolution layers are translation-equivariant instead of translation-invariant. The equivariance allows CNN to generalize edge, texture, and shape detection in different locations. Pooling passes over the max value in its receptive field regardless of its spatial position brings the ability of translation-invariance to the CNN.

[6] The receptive field is a hyperparameter defined as the spatial extent of connectivity where each neuron in the layer is connected to only a local region of the input data.

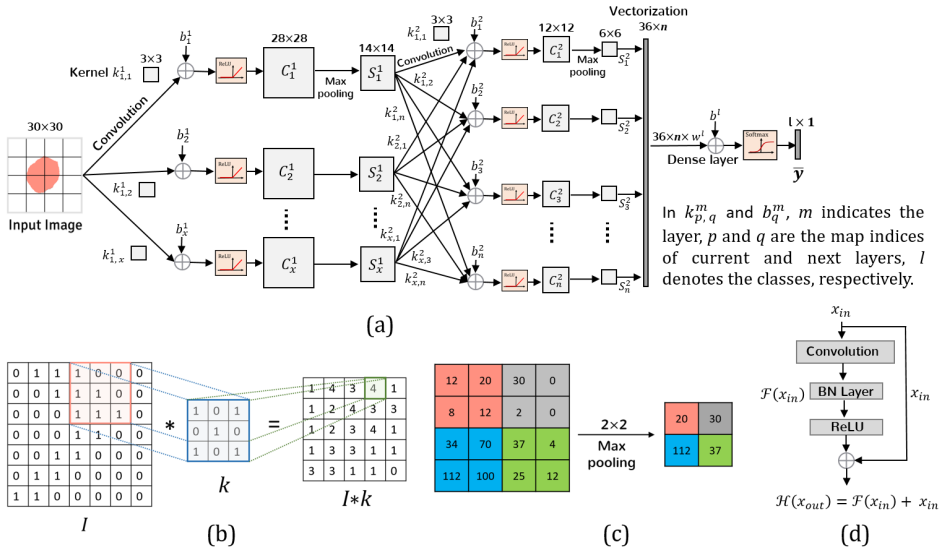


Figure 2.5 Building blocks of a CNN architecture. (a) A typical CNN architecture consists of several convolutional layers and pooling layers on top of the dense layers, (b) convolution filter $k_{3\times3}$ that convolves throughout the whole input image $I_{7\times7}$, (c) an example of max-pooling operation using a filter of size 2×2 with the stride of 2 applied on the input, and (d) an example of residual mapping with one convolution block consisting of a convolutional layer, a batch normalization (BN) layer, and a ReLU activation function.

channel of a convolutional layer for a kernel k and bias b can be formalized as follows:

$$\text{conv}(I,)_{xy} = f\left(b + \sum_{i=1}^{m}\sum_{j=1}^{m} k_{ij} \cdot I_{x+i-1,y+j-1}\right), \tag{2.14}$$

The dimensions of the resulting feature maps are controlled by three hyperparameters which are required to be specified before the convolution step is performed such as depth, stride, and zero-padding. First, the depth corresponds to the number of kernels use for the convolution operation. In the network shown in Fig. 2.5(a), the convolution operation is performed on the input image using x kernels, thus producing a depth of x different representation maps. Second, the stride is the number of pixels by which the filter shifts over the input image in each step to compute the next pixel in the result. It specifies the overlap between individual output pixels. For instance, when the stride is 1, the filters shift by one pixel at a time, and when it is 2, the filters shift over 2 pixels in each step. Third, convolution operation using a kernel larger than 1×1 reduces the output dimension of an image. Therefore, padding is often desired to keep output spatial dimensions the same as input where the image is sufficiently padded with zeros at the borders. For instance, a kernel of size $m \times m$ is used then a zero padding of size $\frac{m-1}{2}$ is added to the border of the input image. For a given input

image of hight or width I_y with a stride S and padding P, the dimensions of the output of any given convolutional layer C can be obtained as:

$$C = \frac{I_y - m + 2P}{S} + 1 \ , \tag{2.15}$$

Pooling Layer

The pooling layer is often synonymized as a subsampling (or downsampling) layer which usually follows the convolutional layers. Pooling layer can be seen as a regularization layer that controls the overfitting of a network. It progressively reduces the spatial dimensions of the given representation maps and thus leading to less computational heads for the next layers.

There are several operations to implement pooling such as max-pooling, L_2 norm pooling and global average pooling. However, max-pooling, which find the maximum value of the input patch, is the most popular for the classification tasks. Max-pooling is often applied using filters of size 2×2 and a stride of 2 at every depth slice. An example of a max-pooling operation using a filter of size 2×2 with the stride of 2 applied on the input is shown in Fig. 2.5(c).

Dense layer

The dense (or fully-connected) layers follow the same connectivity as in an MLP where each neuron of the input layer is connected to every neuron in the succeeding layers. The convolution layers are modeled to extract the discriminative representations (as a feature extractor), whereas the dense layers are modeled for classifying the objects in their respective classes. Therefore, this connectivity is different from the local connection style employed in the convolutional layers. The dense layers are implemented by structuring (or flattening) the input feature maps into a vector, followed by vector-matrix multiplication, then a bias term is added to it. Finally, a transfer function is applied to induce the non-linearity as follows:

$$h^l = f\left(b^l + W^l\, h^{l-1}\right) , \tag{2.16}$$

where h^l is the output feature vector of the layer l which is obtained by flattening the input feature maps h^{l-1} of the $l-1$ layer; W^l, b^l, and $f(\cdot)$ are respectively the weight matrix, the bias term, and the transfer function.

Residual connections

Increasing network depth imposes challenges from the optimization perspective and also regarding the overall performance of the CNN. With increasing depth of the network, the accuracy saturates and starts to degrade rapidly due to the vanishing gradient problem. To overcome such challenges, residual (or skip) connections can be added to the network topology.

The key concept of residual mapping is to introduce modularity [7] in the deep networks where identical mapping is performed by summing the input of one layer to the output of at least one skipped layer [29]. The residual mapping is based on the approximation of the residual function instead of the original one directly from a convolutional layer $\mathcal{H}(\cdot)$, and is expressed as:

$$\mathcal{H}(x_{out}) = x_{in} + \mathcal{F}(x_{in}, \{W_k\}) \ , \tag{2.17}$$

where, x_{in} and x_{out} are its input and output; $\mathcal{F}(\cdot)$ is a residual mapping associated with a set of parameters $\{W_k\}$ where $k \geq 1$, means skipping at least 1 convolutional block, consisting of a convolutional layer, a batch normalization (BN) layer and rectified linear unit (ReLU) activation layer (Figure 2.5(d)). From learned feature weights sharing perspective, residual connection enables feature reuse at no extra parameters and computational complexity. In addition, it allows the gradient to flow through them during the backward propagation easily. The residual connections are used in Papers II and IV.

2.7. Supervised Optimization of Deep Neural Networks

The performance of DNN is optimized in conjunction with minimizing the objective (or loss) function, which is challenging since the loss function is high dimensional and non-convex. Provided this, it is reasonable to employ iterative-based optimization algorithms for finding a parameter configuration to minimize the objective function $\mathcal{L}$.

Most of the iterative algorithms are based on the GD method. The GD minimizes the objective function by updating the parameters ($\boldsymbol{\theta} = [w, b]$) in the negative gradient direction of the objective function $\nabla_\theta \mathcal{L}(\boldsymbol{\theta})$. The step size is determined by the learning rate η. It can operate in the batch, stochastic, and mini-batch learning modes.

Batch Gradient Descent (BGD) updates the parameters after computing gradients of all the samples at once. It can be very slow and is not feasible for large datasets.

$$\boldsymbol{\theta} = \boldsymbol{\theta} - \eta \cdot \nabla_{\boldsymbol{\theta}} \mathcal{L}(\boldsymbol{\theta}) \ , \tag{2.18}$$

Stochastic Gradient Descent (SGD) updates the parameter for each training sample x_i and label y_i by performing one update at a time. Although it is faster than BGD, its frequent updates lead to high variance in the parameters. These fluctuations of parameters overshoot the loss function to different suboptimal minima, and thus, ultimately leads to unstable convergence.

$$\boldsymbol{\theta} = \boldsymbol{\theta} - \eta \cdot \nabla_{\boldsymbol{\theta}} \mathcal{L}(\boldsymbol{\theta}; \, x_i; \, y_i) \ , \tag{2.19}$$

[7] Modularity refers to as a small network that can be repeated to increase the depth of the network.

Mini-Batch Gradient Descent (MBGD) is an improved variant of BGD and SGD optimizers. It updates the parameters on a smaller batch of n training samples. In comparison to SGD, it converges smoothly due to reduced variance in the parameters. Typically, the batch sizes range between 32 and 256. The MBGD learning algorithm is used in Papers II and III.

$$\boldsymbol{\theta} = \boldsymbol{\theta} - \eta \cdot \nabla_{\boldsymbol{\theta}} \mathcal{L}(\boldsymbol{\theta};\ x_{(i:i+n)}; y_{(i:i+n)}) \ , \tag{2.20}$$

Momentum is an adaptive version of MBGD [30]. It accelerates MBGD by softening its convergence in irrelevant directions. Since MBGD are prone to stick in saddle points[8], Momentum navigates it along the relevant direction by using an average gradient over the previous steps.

$$\begin{aligned} v_t &= \gamma v_{t-1} + \eta \nabla_{\boldsymbol{\theta}} \mathcal{L}(\boldsymbol{\theta}) \ , \\ \boldsymbol{\theta} &= \boldsymbol{\theta} - v_t \ , \end{aligned} \tag{2.21}$$

where γ is the momentum term, v_t, and v_{t-1} are respectively the current and previous updates to the parameters.

Adaptive Gradient (Adagrad) is an adaptive optimization algorithm [31], which is best-suited for the sparse data. It updates the learning rate by scheduling a priority for each parameter. It means that the infrequent parameters are prioritized with larger updates whereas frequent parameters are assigned with a priority of small updates. The update for each parameter, $\boldsymbol{\theta}_i$, with a different learning rate at step k is computed as:

$$\boldsymbol{\theta}_{k+1,i} = \boldsymbol{\theta}_k - \frac{\eta}{\sqrt{G_{k,i} + \epsilon}} \nabla_{\boldsymbol{\theta}} \mathcal{L}(\boldsymbol{\theta}_{k,i}) \ , \tag{2.22}$$

where $G_{k,i}$ is a diagonal matrix with each diagonal entry as the sum squared of the gradients of $\boldsymbol{\theta}_i$ up to step k and ϵ is a small value to prevent division by zero.

Root mean square propagation (RMSProp) is also an adaptive optimization algorithm [32], eliminating the problem of gradient accumulation[9]. It overcomes the issue of radical diminishing of learning rates raised by Adagrad. It updates the parameters iteratively with a running average of squares of previous gradients. It prevents gradients from exploding by decreasing the step size for larger gradients, and from vanishing by increasing the step size for small gradients. The average squared gradient, $E[g^2]_k$, for step k is defined on the average at step $k-1$ and the current gradient as:

$$\begin{aligned} E[g^2]_k &= 0.9E[g^2]_{k-1} + 0.1g_k^2 \ , \\ \boldsymbol{\theta}_{k+1} &= \boldsymbol{\theta}_k - \frac{\eta}{\sqrt{E[g^2]_k + \epsilon}} g_k \ , \end{aligned} \tag{2.23}$$

[8] Saddle point is a point where one dimension slopes up while another slope down, usually surrounded by a plateau of about equal error. Regardless of the direction, it is difficult for the non-adaptive variants of GD to converge since surrounded gradients are nearly zero.

[9] Accumulation means running summation. The gradients are a running summation of every new batch which is computed after propagating backward on one batch at a time.

where $g_k = \nabla_{\boldsymbol{\theta_k}} \mathcal{L}(\boldsymbol{\theta}_k)$, and ϵ is a small value to prevent division by zero. There are several other variations of adaptive optimizier that are not discussed but a comprehensive overview can be easily found in [33]. The RMSProp learning algorithm is used in Papers I, IV, and V.

2.8. Regularization of Deep Neural Networks

Overfitting is one of the challenges that is often encountered when training DNN. It occurs when the parameters of a model are optimized well without capturing the underlying representations of the data. It implies that certain complexities in the model may restrict the model to generalize well even though it fits well with the training data. Regularization is a technique which reduces overfitting and variance in the model by penalizing its complexity. It is added to the model so that it can fit adequately to the training data but at the same time it can generalize better to unseen data.

Regularization can directly be added as a penalty term to the loss function that penalizes the parameter outliers in the model, i.e., large weights. This kind of regularization is often referred to as parameter norms or weight penalty terms. A slightly different approach is to modify the network by dropping its parameters randomly while training which can be achieved using "dropout layers". In this thesis weight decay, data augmentation and dropout have been used to prevent overtraining and to improve generalization.

Dataset augmentation

Limited amounts of annotated data pose severe challenges while training a supervised DNN model. Given that CNN has a large number of parameters and hyperparameters[10] to be optimized, overfitting is thus highly probable to occur when training a model with very few samples. Availability of more data can certainly improve the overall performance of a model. One possible solution is to artificially augment the dataset by generating a moderate amount of new yet correlated training samples.

The choice of selecting an augmentation technique is often problem specific[11]. For instance, pulmonary nodules on CT scans pose a great variability regarding contextual surrounding, shape, size, and orientation. Generic augmentation techniques such as rotation and translation can be employed to augment the training set for a classification problem. In such a way, the model can learn rotational- and translational-invariant features, and thus improve the overall performance of the classifier.

[10] Hyperparameters represent the configurable values used when building a network such as filter sizes, learning rate, dropout, gradient clipping threshold, etc; whereas parameters constitute the learnt values (weights) obtained while optimizing the loss function.

[11] An interesting fact: There are 48 unique lossless permutations of 3D images possible as opposed to only 8 for 2D images. `https://en.wikipedia.org/wiki/Octahedral_symmetry#The_isometries_of_the_cube`

Cross-validation

In supervised learning, the performance of a model is often measured by holding out a validation set from the training data. Since there is typically a shortage of data to train a model, removing a part of the data for validation poses a problem of underfitting. In case of data scarcity, cross-validation is an efficient statistical method for evaluating the performance of a model which also helps in the overall generalizability of a model.

In a typical k-fold cross-validation scheme, the data is randomly split into F_k equally sized folds (e.g., $k = 5$). Subsequently, k iterations of training and validation are performed in such a way that within each iteration a different fold is used for validation while the remaining F_{k-1} folds are used for training. In such a way, each data point in the training and validation sets cross-over in successive rounds and gets a chance of being validated against themselves. The final results are determined by either taking a mean or median of measures over the k folds. However, it is also recommended to perform the final evaluation using a completely unseen test set since cross-validation might bias the model to generalize to both the training and validation data due to cross-over sampling. The 5-fold cross-validation scheme is used in Papers I, II, IV, and V.

Dropout

Dropout regularizes a model by reducing the interdependent learning amongst the neurons and eventually prevents it from overfitting. Dropout can be seen as a version of bagging where some neurons are randomly dropped out at every iteration so that they will not interact with the network. It implies that the weights for those dropped neurons are not updated, and they do not affect the optimization of other neurons in the network.

In such a way, a sparse network composed of several networks is developed where each network is trained with a single sample. Such transformation of a network into an ensemble hugely decreases the possibility of overfitting. Since the influence of individual neurons on learning is averaged, it helps a network to generalize better and also increases accuracy. The Dropout technique is used in Papers I, II, IV, and V.

Batch Size

Training a CNN with GD optimizers on relatively larger batch sizes[12] can influence the convergence to sharp minimizers[13] (Fig. 2.3(c)), and thus, potentially affect the network generalizability. On the other hand, training with smaller batches can lead the convergence to flat minimizers[13] (Fig. 2.3(c)) due to the inherent noise in the gradient estimation. In either case, the network will not

[12]Batch or mini-batch size is referred to as the number of training samples in one forward/backward propagation.

[13]Flat minimizers is defined as the size of the connected region around the minimum where the training loss is relatively similar. Consider the error as a one-dimensional curve, a minimum is flat if there is a wide region around it with roughly the same error; otherwise, it's sharp.

be possibly able to converge to an optimal set of parameters and will result in poorer generalization. Although determining a batch size is a somewhat empirical practice, it is still an important hyperparameter which also influences the overall generalizability of a network [34].

Batch normalization

Variations in the parameters from each layer (i.e., internal covariate shift) slow down the network training with saturated activations and small learning rate. This can adversely affect the training of the CNN with a risk of poor generalization performance [16]. Lately, batch normalization (BN) has enabled the CNN to learn faster with a better generalization of the network and overcome the issue of internal covariate shift[14]. While training with BN, each feature map computed by a linear operation (e.g., convolution) is normalized separately over the mini-batch[12] to have a mean μ of zero and variance σ^2 of 1. For example, a layer with an input $\mathbb{X} = (\mathrm{x}_1, \cdots, \mathrm{x}_m)$, where *m* is the total number of feature maps computed after applying a linear operation. Each x_n is formed by all the corresponding feature maps of the candidates in the mini-batch (e.g., 128). The BN for n^{th} feature map can be expressed as:

$$\hat{x}_n = \frac{x_n - \mu(x_n)}{\sqrt{\sigma^2 [x_n]}} \ , \tag{2.24}$$

However, just simply normalizing the feature map can constrain the representation capabilities of the network. Therefore, a pair of learning parameters (learned along with the original model parameters) for scaling by γ_n and shift by β_n is applied to the normalized feature map $\hat{x}_n$ as:

$$y_n = \gamma_n \hat{x}_n + \beta_n \ , \tag{2.25}$$

By employing BN, the network converges much faster and also improves the overall generalization of the network. Although BN reduces the strong dependence on initialization, it is still often beneficial with proper initialization of weights. The batch normalization technique is used in Papers II, IV, and V.

Weight penalty

Penalizing the weights while training is one of the conventional techniques to regularize a network. With an implicit assumption that a model with small weights is somehow simpler than the one with large weights, the penalties try to reduce the complexity of a model by keeping the weights small. The structural risk (or loss) function using the regularization term, and empirical risk term from Equation 2.7 can be expressed as:

$$\mathcal{L}(\boldsymbol{\theta}) = \frac{1}{n} \sum_{i=1}^{n} \left(y_i, f(\mathbf{x}_i, \boldsymbol{\theta}) \right) + \lambda \, \Phi(\boldsymbol{\theta}) \ , \tag{2.26}$$

[14]Internal covariate shift refers to as the change in the distribution of network activations due to the change in network parameters during training.

where λ is the regularization hyperparameter to control the influence of regularization term $\Phi(\boldsymbol{\theta})$. Larger values of λ imply more regularization, i.e., smaller values for the model parameters $\boldsymbol{\theta}$. The regularization terms can either be L_2 norm (Ridge Regression) or L_1 norm (Lasso Regression). The L_2 norm adds a penalty proportional to the squared magnitude of each weight as follows:

$$\Phi_{L_2}(\boldsymbol{\theta}) = \frac{1}{2} \sum_j \boldsymbol{\theta}_j^2 \,, \tag{2.27}$$

The L_2 norm penalizes larger weights more (weights are squared) and restricts the parameters to small non-zero values. The L_1 norm penalizes the absolute value of the weights and is defined as:

$$\Phi_{L_1}(\boldsymbol{\theta}) = \sum_j |\boldsymbol{\theta}_j| \,, \tag{2.28}$$

It introduces sparsity in the network by equally penalizing the smaller and larger weights. This sparse property is often helpful when selecting important features. The L_2 norm weight penalty is used in Papers II and III.

Early stopping

In supervised learning, the given set of data points is split into a training set and a test set. The training set is further divided into a training subset and a validation subset. The training subset is used to find the hyperparameters (learning process), and the validation subset is used to tune the parameters (hyperparameter tuning). At any given point of the learning process, the test set is not used for optimizing the hyperparameters. The test set is used for model selection or for accessing the performance of individual model trained using cross-validation schemes. Skipping the test set is not a feasible choice since the algorithm that performed well during the cross-validation does not guarantee a good performance due to the possibility of inheriting noise in the cross-validation set.

When training a model using iterative-based learning algorithms, the performance (e.g., loss function or accuracy) on the training subset cannot be used as an assessment criterion since the model may get tuned to the noise present in the training data. In that case, the validation subset is used for evaluating the performance of a model. However, the error on the validation subset might begin to grow when the network starts overfitting to the data. When the validation error increases for a specified number of iterations in a row, the training is stopped, and the parameters at the minimum of the validation error are returned. This is usually considered as early stopping and implies a similar regularization like L_2 norm regularization. The early stopping is used in Papers II, III, and IV.

Curriculum Learning

Curriculum learning is inspired by the fact that systematically organized learning can lead to better understanding of complex concepts (much like human learning). It helps a model to generalize better by increasing the influence of simple

data-points. In curriculum learning, the core feature to enable is to rank the data-points based on their level of presumed difficulty and then train a model with simple data-points first before gradually progressing to harder data-points.

This strategized learning can be employed by assigning larger weights to the simple data-points in a loss function or by sampling them more frequently. An appropriate curriculum strategy, therefore, both acts to help the learning process and to regularize by giving rise to lower generalization error for the same training error. One such learning strategy is used for training a CNN classifier in Paper I.

2.9. Transfer learning

The concept of transfer learning is one of the extended benefits of DNN-based methods where a model trained on one task can be reused to comprehend the problems associated with another related task. On a conceptual level, transfer learning is intrinsically connected to the idea of generalization. The primary distinction is that transfer learning is often used for transferring knowledge across tasks instead of generalizing within a specific task. More specifically, transfer learning uses the representations learned from tasks for which a lot of labeled data is available compared to the settings with only little-labeled data.

To leverage the transfer learning perspective, a pre-trained DNN model can be employed either as for feature extraction or for fine-tuning it on a new task. As a feature extractor, a pre-trained model with optimized parameters is used to extract representations for a new set of data points. Then, the dense layers are trained using those representations for the classification task. In a fine-tuning process, the parameters of a pre-trained model are re-optimized by continuing the backpropagation using a small learning rate to refine the parameters for a new dataset. It is possible to fine-tune each layer or to fine-tune later layers of a pre-trained CNN. It is often recommended to fine-tune later layers since the earlier layer extracts more generic representations that could be useful to many tasks, later layers become progressively more problem-specific to the details of the classes in the respective dataset. One such aspect of transfer learning is shown in Paper II and V.

3. CAD FOR PCD ANALYSIS IN TEM IMAGES

Manually diagnosing primary ciliary dyskinesia (PCD) using transmission electron microscopy (TEM) is time-consuming, subjective, and monotonous. Automation of the process is thus highly desirable to assist pathologists. However, developing an automated workflow to mimic the manual diagnostic procedure imposes several challenges regarding image acquisition and analysis. Amongst many of such challenges, this chapter explicitly focuses on the problems associated with the detection and denoising.

To improve the performance of an automated PCD analysis workflow, two CNN–based methods are developed to 1) classify cilia and non-cilia instances in low-magnification TEM images (Paper I), and 2) denoise short exposure high-magnification TEM images for enhanced ultrastructural analysis (Paper II). This chapter summarizes the background, material, methods, results, and contributions presented in the appended publications (A and B).

3.1. Overview

Cilia are hair-like structures protruding from cells surface. Dysfunctional cilia are often associated with a genetic disorder –*Primary Ciliary Dyskinesia* (PCD), which can lead to pulmonary infections, reduced female fertility, and infertility in males [35, 36]. Early and accurate diagnosis is highly desirable to control the progression. In 1976, Afzelius reported that the ultrastructural defects of cilia lead to immotile cilia and are primarily responsible for immotile cilia syndrome (ICS) [37]. Later, the term ICS was replaced by PCD to distinguish genetic ciliary defects (primary) from defects due to viral respiratory tract infections or exposure to toxic agents (secondary) [35, 38, 39].

The prevalence of PCD is rather difficult to estimate since most of the patients often remain undiagnosed due to nonspecific symptoms, insufficient knowledge of the disease, and limited diagnostic facilities [40]. The estimated prevalence varies between one in 2,000-40,000 [41, 42]. PCD diagnosis is challenging since there is specifically no such diagnostic test which is accurate enough to be used as a stand-alone test. European Consensus guidelines recommend combining tests, including nasal nitric oxide, high-speed video microscopy, TEM, and genetic culture testing [43–45]. Such tests are often expensive, requiring highly-skilled staff and technically advanced equipment, which limits them to

highly specialized centers [40]. This work is explicitly centered on facilitating the manual diagnostic procedure of PCD analysis using TEM.

To diagnose the disorder, a trained pathologist analyzes the morphological appearances of at least 50 perfectly perpendicularly cut cilia ($\sim$220–250 nm in diameter) in high-resolution TEM images [46]. TEM can reveal morphological structures up to a resolution of 1 nm; however, it is an expensive and technically complex imaging technique [47, 48]. In a typical clinical setting, pathologists steer at different magnification levels to identify diagnostic-relevant cilia instances. Manually locating and analyzing such nanostructures in the large search space a tissue sample constitutes is monotonous, and time-consuming (this could take up to two hours per patient). Automated image acquisition and analysis are thus vital to improve the PCD analysis using TEM.

3.2. False Positive Reduction in Low Magnification TEM Images

Diagnostic quality cilia instances are rare, very small, and are often unevenly spread throughout the sample in the form of clusters. To cover a large search space of tissue sample in a reasonable time, it is important to use as low-magnification as possible. This entails the analysis of LM images [I]. Automatically locating potential regions of interest at low magnification, and acquiring high-magnification images only at selected locations, is therefore highly beneficial.

Automated detection of cilia in low-magnification TEM images is challenging due to the heterogenic quality among cilia instances and their similar characteristics with non-cilia candidates. In a 4K$\times$4K LM image with a FOV of $\sim 60^{'}$um, a cilium instance is of about 20 pixels in diameter, and thus its characteristics can be barely resolved to compute discriminative features [I]. Previously, a template matching method to detect cilia candidates in low-magnification TEM images [46] has been proposed. This method achieves considerable detection performance; however, it also introduces a large number of FPs. The template matching-based methods often depend on a local cross-correlation that is relatively sensitive to noise and therefore potentially detects a substantial fraction of FPs [49, 50].

While aiming at locating highly populated regions of diagnostic-quality cilia for further high magnification image acquisition and analysis, it is imperative that such regions are not misled by a large number of FPs. Given this, it is crucial to employ an FP reduction module in the automated workflow to improve the overall detection performance. Lately, increased computational power and availability of a large amount of data has increased the applicability of DNN-based methods in the biomedical image analysis field [51–54]. The capacity of CNN to encode the discriminative representations in a supervised learning regime makes them efficacious for automated detection of structures. Motivated by that, a CNN classifier is developed with a particular focus on reducing the number of FPs

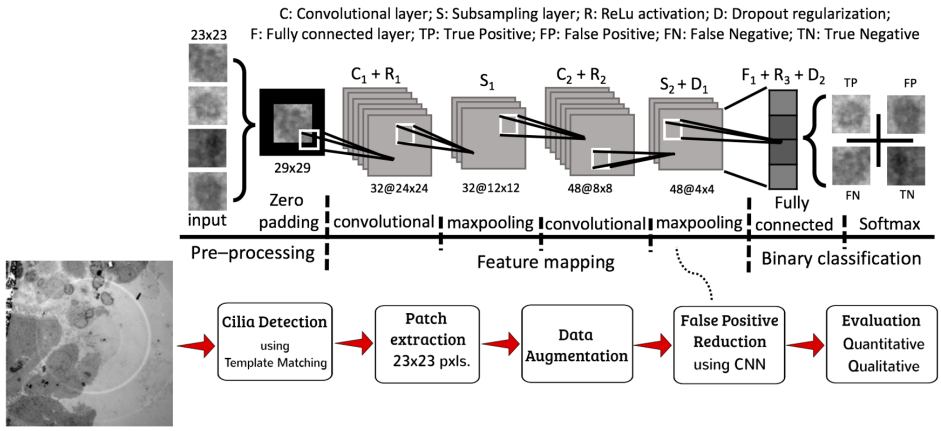


Figure 3.1 An overview of the overall workflow consisting of the CNN model.

and is integrated with the existing template matching-based candidate-screening module.

Method

The schematic illustration of the overall detection workflow is shown in Fig. 3.1. It consists of two stages: (1) template matching to detect the plausible cilia candidates, and (2) further FP reduction using a 2D CNN model. This chapter only focuses on the FP reduction (Publication A) whereas the former part of the workflow is comprehensively explicated in [46].

Template matching based on normalized cross-correlation (NCC) and a customized synthetic template are used to detect the initial cilia candidates. Subsequently, patches of 23×23 pixels are extracted for the initially detected candidates as the input for the CNN classifier. The patch contains a cilium ($\sim$ 19–20 pixels in diameter) and some local background around it (~ 3 pixels) to include sufficient context information.

Given the complexity of cilia detection in the LM images, it is worthwhile to employ an organized training strategy where the complex candidates are introduced systematically. Such organized training is typically associated with curriculum learning (discussed in Chapter 2). To infer curriculum learning, a training set of cilia, as well as non-cilia candidates, is extracted from the training image. The training set includes all 136 true cilia instances regardless of their NCC values. In addition, it also includes 272 non-cilia candidates from different NCC levels to represent non-cilia candidates with high similarity to good cilia (136 randomly chosen non-cilia objects with NCC values ≥ 0.5) as well as non-cilia candidates less similar to cilia (136 randomly chosen objects with NCC threshold values between 0.2 and 0.5).

An imbalanced dataset can mislead the optimization algorithm to converge to a local minimum, wherein the predictions can be skewed towards the candidates of the majority class. Therefore, candidates from both classes (i.e., cilia and

non-cilia) are augmented to overcome the overfitting issues. To balance the sets, horizontal flipping is applied only to the cilia candidates. After that, seven random angular rotations (0–360°), six random scalings within ±10% range and five random shearings within 5% range in both x- and y- directions are applied, resulting in 1050 augmented variations for each candidate.

The CNN classifier is an adapted version of the LeNet model [55]. It consists of two convolutional layers and two max-pooling layers, as shown in Fig. 3.1. Firstly, the input patches are padded with a three-pixel thick frame of zeros to keep the spatial sizes of the patches identical after the first convolutional, as well as to keep the border information up to the last convolutional layer. The first convolution layer generates 32 feature maps using 6×6 convolutions. The second convolution layer generates 48 feature maps using 5×5 convolutions. The max-pooling layer downsamples the feature maps by selecting the maximum feature response in windows of size 2×2. The fully connected layer consists of 20 neurons and is followed by a Softmax layer to predict the final probability distribution of the input candidate. Each convolution layer and fully connected layer is followed by a ReLU [56] non-linear activation.

Before the training, the patches are normalized by subtracting the mean and dividing by the standard deviation. The weights are initialized using Glorot normal distribution [23], and the biases are initialized with zeros. The weights are adaptively updated in mini-batches of 128 candidates using the RMSProp optimizer [32]. The training runs for 50 epochs in a five-fold cross-validation scheme with a learning rate of 0.001. A dropout [57] layer with a probability of 0.5 is implemented on the output of the last pooling layer and the output of the fully connected layer. The error loss is measured using the softmax loss function. The 2D CNN is implemented using Theano backend in Keras.

Material and Evaluation Criteria

Two low-magnification (LM) TEM images from different patients, consisting of ca. 200 cilia instances, are used for training and testing purposes. Both images are acquired with an FEI Tecnai G2 F20 TEM, and a bottom mounted FEI Eagle 4K×4K HR CCD camera, resulting in 16-bit grayscale TIFF images of size 4096×4096 pixels. For each LM image field, a set of mid-magnification (MM) images are acquired, where the true cilia candidates of diagnostic quality are manually marked by an expert pathologist (author AD in Paper I). Some examples of extracted patches of marked cilia candidates are shown in Fig. 3.2. The FOV for a MM (2900×) image is 15.2 µm and 60.6 µm for a LM (690×) image. The performance is evaluated using AUC_{PR} and F-score (discussed in Section 1).

Results and Discussion

The performance of CNN is investigated at different NCC threshold levels (0.2 - 0.5) determined by the template matching method. The F-score curves for

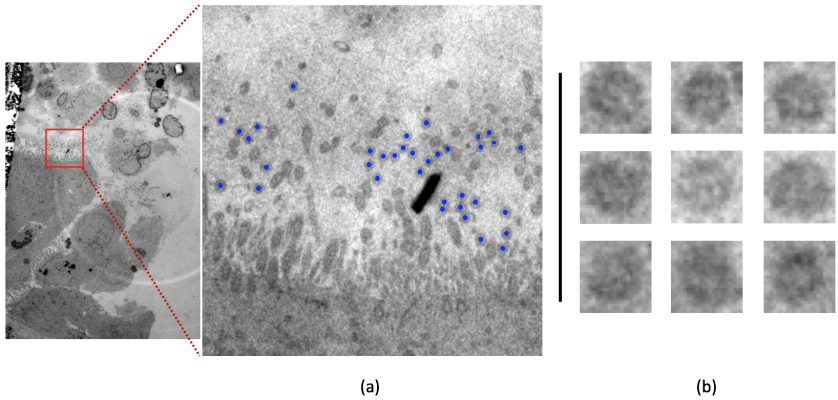


Figure 3.2 Low magnification TEM image of 4096×4096 pixels utilized for training purpose with the magnified view of 350×350 pixel bounding box (marked in red) with indicated ground truth marked by an expert pathologist. Here, cilia candidates marked with blue dots are of the suitable quality. (b) Some examples of patches extracted by the previously reported method [46], the first and second rows contain TP whereas patches in the third row are FP.

the detection workflow with and without using CNN classifier at different NCC threshold levels are shown in Fig. 3.3.

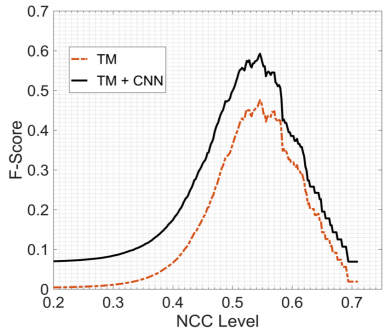
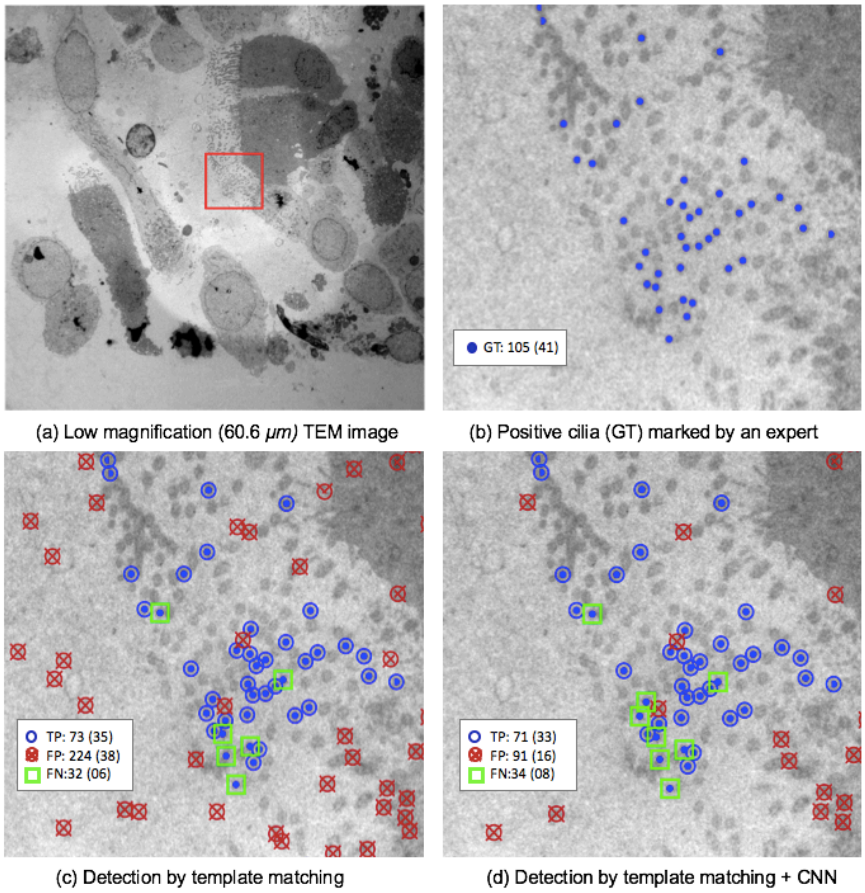


Figure 3.3 F-score curves, for the test image, showing the improvement in the performance by a CNN classifier with template matching method [46] at different NCC threshold levels.

Although the CNN classifier considerably reduces the FPs at all these NCC values, it is not practically suitable as lowering the NCC threshold increases the number of candidates to analyze tremendously while only rather few additional true candidates are detected. Hence, the NCC threshold value is set to 0.5, resulting in an improved F-Score of 0.59 compared to 0.47 for the template matching method. The CNN also significantly improves the AUC_{PR} to 0.82 and 0.71 compared to the previous AUC_{PR} of 0.48 and 0.42 for the template



(a) Low magnification (60.6 *μm)* TEM image (b) Positive cilia (GT) marked by an expert

(c) Detection by template matching (d) Detection by template matching + CNN

Figure 3.4 Illustration of cilia detection results. (a) The 4096×4096 test image, (b) a 650×650 example subregion of the test image, (c) same sub-region after initial template matching method, and (d) after proposed CNN classifier. The numbers are given for the whole image and for the ROI is in parenthesis. Here, blue circles, red crossed circles, and green squares represent the TP, FP, and FN, respectively.

matching method, for the training and test images, respectively, at an NCC threshold level of 0.5.

Detection results of the proposed CNN model on an ROI of 650×650 pixels, in the LM test image, at an NCC level of 0.5, are shown in Fig. 3.4. It shows the detection results of the initial candidate detection step template matching method [46] (Fig. 3.4(c)) and the improved results achieved by incorporating the CNN classifier as an FP reduction step (Fig. 3.4(d)). In these images, the blue circles, red crossed circles, and green squares represent the candidates that have been correctly detected (TP), the candidates that have been erroneously detected as cilia (FP), and the cilia that were missed with respect to the manually ascertained ground truth delineations and initial detection step (FN), respectively.

These results show the potential of proposed CNN model for cilia detection in low-magnification TEM images.

3.3. Denoising of Short Exposure High Magnification TEM Images

Both manual and automated analysis of TEM images are negatively influenced by many imaging artifacts such as non-optimal microscope alignment and focusing, as well as motion artifacts due to sample drift and vibrations (Publication B). Besides that, acquiring high magnification images can potentially damage the sample due to contentious exposure to electrons for a relatively longer period. The imaging artifacts can be reduced by decreasing the electron dose and acquisition time. However, this results in images with more noise and thereby questing for denoising as a potential preprocessing step for improved analysis.

In the past decades, several denoising methods have been proposed to improve the quality of the images [58–61]. Although the traditional methods [62–68] have shown promising performance on image denoising task, these methods typically involve a complex optimization problem during the testing stage and requires manual selection of parameters [58, 69]. To overcome these challenges, several learning-based methods using CNN have been proposed [58, 69–73]. The most significant difference between learning-based methods and traditional methods is that they learn parameters for image restoration directly from training data rather than relying on image priors [69]. However, most of these methods are carefully designed only for a certain type of noise, i.e., Gaussian noise, and thereby limiting their potential inference for the imaging devices with mixed noise distribution.

The noise induced by TEM is non-additive and signal-dependent which can be modeled by a mixed Poisson-Gaussian (PG) distribution [74, 75]. Acknowledging the superior performance of CNN in denoising, a novel multi-stream CNN framework is developed to denoise the short exposure high magnification TEM images. The idea of employing a multi-stream architecture is inspired by the adequate performance of the approaches proposed in [58, 72]. Recently, such ensemble learning for denoising have also been proposed for fluorescence microscopy where the outputs from five pre-trained CNN models are cascaded to obtain the final denoised image [76].

Method

The architectural view of developed multi-stream CNN framework is shown in Fig. 3.5. The training of both streams is performed using the contextual information spread over patches of 128×128 pixels. The patches are extracted with an overlapping stride of 16 pixels. It is often the case that the information in a small patch is not sufficient to preserve the structural information. Although the training is performed using patches of size 128×128 pixels, the trained CNN framework can be used for the arbitrary size of patches during the testing stage.

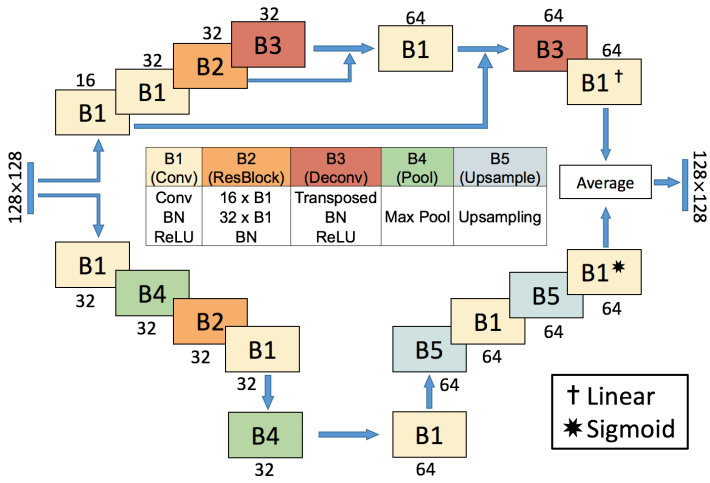


Figure 3.5 The two-stream DCNN architecture. The sizes of the output feature maps of each block are shown on top of each block and generated using 3×3 *convolutions. The last* 1×1 *convolution blocks of each stream use linear and sigmoid activations, respectively, instead of ReLU.*

The first stream consists of four convolution blocks, two transposed convolution blocks and one residual block. The convolution block encodes the image representations while removing the noise, whereas the transposed convolution block decodes these representations to restore the noise-free image content. The residual block contains two convolution blocks. The BN [77] layer is used as regularization before ReLU [56] activation to deal with internal covariate shift. To elevate the training performance, skip connections are used and followed by a BN layer. The second stream consists of four convolution blocks, two up-sampling blocks, two max-pooling layers, and one residual block.

Before the training, the patches are normalized to the range $[0, 1]$. The training is performed averaging the outputs of both streams. The weights are initialized using Glorot normal distribution [23], and the biases are initialized with zeros. The weights are adaptively updated in mini-batches of 16 patches using SGD [78] optimizer. The training continues for 15 epochs in a five-fold cross-validation scheme. The initial learning rate is set to 0.001 and reduced to 1/10 of the current value after every epoch. The error loss is measured using both MSE and a binary cross-entropy function. The multi-stream CNN is implemented using Tensorflow backend in Keras.

Material and Evaluation Criteria

A series of 100 noisy short exposure (2 ms) images, captured at the same spatial location in a cell section sample (FOV = 2000 nm) are used for the training and testing purposes. All images are of size 2048×2048 pixels and acquired

with the low-voltage MiniTEM™[1] system. A low-noise ground-truth image is obtained by registering each short exposure image to the first image of the series using rigid registration, followed by aggregating the information by computing the pixel-wise median value, illustrated in Fig. 3.6. The training of the CNN model is performed using 10 images. These 10 images are also used for explorative parameter tuning of three other classical methods (discussed in Paper II). The remaining 90 images are used for the evaluation and comparison of each method. Apart from evaluating the performances on denoising single images, the performance of each method is also evaluated for two additional denoising strategies 1) *denoising of 5 aggregated short exposure images*, and 2) *aggregation of 5 denoised short exposure images*.

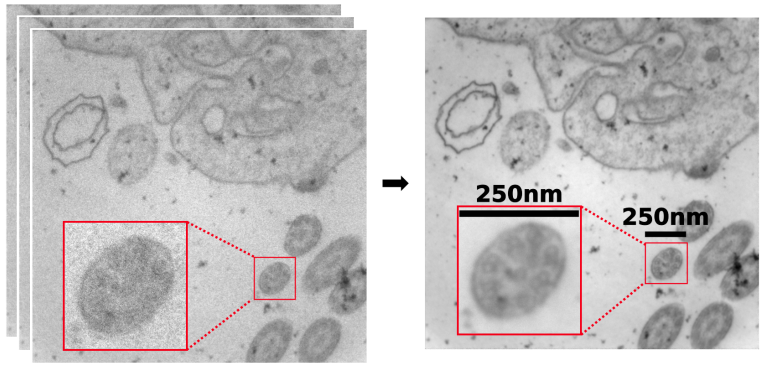


Figure 3.6 **Left***: Short exposure TEM image (*2048×2048 *pixels) from a series of 100 images.* **Right***: Ground truth created by co-registration and aggregation of the stack to the left. The two insets show magnified views (*250×250 *pixels) of one cilium.*

The performance is evaluated using the peak-signal-to-noise ratio (PSNR) and structural similarity index measure (SSIM) [79]. The PSNR is a ratio (in decibels) between the maximum possible value of a signal and the power of distorting noise that affects the quality of its representation. As indicated in [80], different levels of degradations applied to the same image can yield the same PSNR. As SSIM is proposed with the aim to compare structural changes in images imitating what the human visual system does, this measure is considered a more reliable measure of visual similarity of images. Typically, the higher values of both PSNR and SSIM correspond to better image quality.

To validate the level of agreement between the quantitative results and visual (qualitative) results, a subjective visual evaluation conducting a two-step voting process by six of the authors is performed. In the first step, involving only the classical methods, the authors rated the results (1^{st}, 2^{nd}, and 3^{rd} best) on the cilium sub-image produced by each of the methods with different parameter settings. The displayed seven images spanned a parameter range centered around the maximal SSIM for that method. The procedure was repeated for the above

[1] Vironova AB, Stockholm, Sweden

mentioned two strategies of aggregating 5 short exposure images. The second step involves all four methods. The images resulting from the two aggregation strategies utilizing the tuned parameter settings as decided in the first Step, together with the CNN results were displayed (random, unknown order) and the authors rated them again (as the 1^{st}, 2^{nd}, and 3^{rd} best).

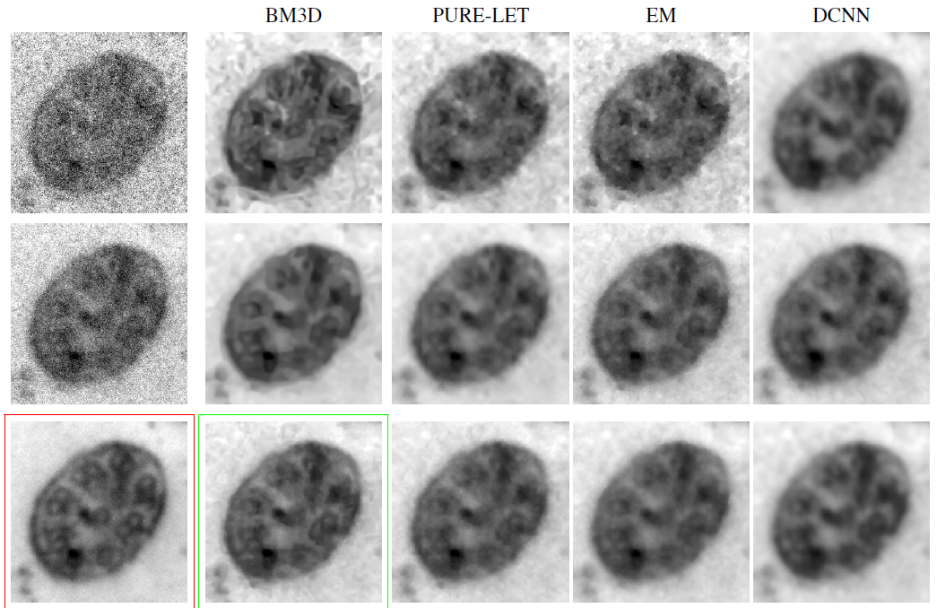


Figure 3.7 Noisy and denoised close ups of a cilium instance obtained with the considered methods. **Top***: Denoising of a single image.* **Middle***: Denoising of 5 aggregated noisy images.* **Bottom***: Aggregation of 5 denoised single images. The red frame (bottom left) indicates the ground truth for single noisy images. The green frame indicates the best ranked image in the two-step visual voting process.*

Results and Discussion

Three classical methods suited for Gaussian and PG noise: a block matching (BM3D) [62], wavelet domain (Pure-LET) [67], and energy minimization (EM) [81] are evaluated and compared with the CNN framework. For the single image denoising task overall 90 images from the test set, the developed CNN framework achieves the highest PSNR as shown in Table 3.1. On the other hand, The EM method marginally performs better than the CNN framework regarding SSIM.

For the *denoising of 5 aggregated short exposure images* strategy, a set of 5 short exposure images are registered and aggregated by the pixel-wise median, resulting in a set of 18 images. A noisy cilium instance from this strategy and the corresponding denoised results obtained with all 4 methods are shown in the middle row of Fig. 3.7. For the *aggregation of 5 denoised short exposure images* strategy, five sequentially acquired short exposure images are denoised, then

registered and aggregated by the pixel-wise median. The corresponding results on the cilium sub-image are shown in the bottom row of Fig. 3.7.

From the quantitative and qualitative results in Table 3.1 and Fig. 3.7, it is clear that denoising improved both single and aggregated short exposure images. Both of the two aggregation strategies improve the results approximately equally well. Based on the visual assessment, the output of the BM3D method from the *aggregation of 5 denoised short exposure images* strategy produces the most appealing result. Overall, CNN gives the highest quantitative scores as confirmed by the average PSNR and the SSIM in Table 3.1. Given that the CNN is trained using the single frame images, it is impressive that the CNN also performs equally well on the other two strategies, thus, showing the transfer learning perceptive of learning-based methods.

Table 3.1 Results on the test data set. Average PSNR and SSIM (± standard deviation) over 90 single images are given in the 1st and 2nd rows. Rows 3 and 4 contain average PSNR and SSIM over 18 aggregated groups of 5 short exposure images followed by denoising. Average PSNR and SSIM over 18 images each obtained by aggregating 5 denoised short exposure images, are given in rows 5 and 6. Best performances are marked in bold.

		Initial	BM3D (σ_{bm})	PURE-LET (σ_{pl})	EM (λ)	DCNN
1	PSNR	22.25	37.39 ± 0.30	37.38 ± 1.09	37.80 ± 0.27	**38.04** ± 0.21
	SSIM	0.019	0.233 ± 0.007	0.219 ± 0.007	**0.255** ± 0.027	0.252 ± 0.002
2	PSNR	27.88	40.45 ± 1.09	40.19 ± 1.06	40.19 ± 0.54	**40.86** ± 0.37
	SSIM	0.037	0.270 ± 0.019	0.263 ± 0.017	0.277 ± 0.017	**0.282** ± 0.011
3	PSNR	22.25	39.65 ± 1.04	40.21 ± 0.48	39.92 ± 0.93	**40.84** ± 0.45
	SSIM	0.019	0.261 ± 0.013	0.265 ± 0.011	0.273 ± 0.021	**0.276** ± 0.009

4. CAD FOR PULMONARY NODULES IN CT IMAGES

The early manifestation[1] of lung impairments, e.g., lung cancer is vital for effective treatment planning and thus considered as key to minimizing the high risk of death. However, manual interpretation of thoracic CT scans for the detection of different sizes of pulmonary nodules is a tedious, labor-intensive, and time-consuming task.

To facilitate the manual interpretation process, two discrete CAD systems as assistive tools are developed for the early manifestation of multiple sizes of nodule candidates in CT scans. The first CAD system (Paper III) is discussed in Section 4.2, and aims to detect pulmonary nodules associated with lung cancer using a combination of classical image analysis and neural network-based methods. The second CAD system (Paper IV) is discussed in Section 4.3, and aims to detect micronodules associated with the fatal and incurable occupational pulmonary disease (silicosis) using a combination of classical image analysis and 3D CNN-based methods. This chapter summarizes the background, material, methods, results, and contributions presented in the appended publications (C and D).

4.1. Overview

Cancer is the leading cause of death around the world, and lung cancer is the second most common cancer, following prostate and breast cancers in men and women, respectively [82]. American Cancer Society estimates that lung cancer accounts for 27% of new cancer cases and 23% of cancer deaths in 2018, i.e., one out of four cancer deaths [82]. With an estimated 275 700 deaths (approximately 20%), lung cancer is the leading cause of all cancer deaths in Europe [83]. The overall 5-year survival rate for men and women in Europe is only 11.2% and 13.9%, respectively [84]. Unfortunately, the majority of cases are diagnosed in the late stages of cancer progression, resulting in ineffective treatment planning and a high mortality rate. Considering that the 5-year survival rate is 56% for early-stage cancer [85], it is worthwhile to detect pulmonary nodules in the early stages.

Smoking is by far the most influential risk factor for lung cancer which is further influenced by the quantity and duration of smoking. Although smoking

[1] In this chapter manifestation refers to detection

accounts for 80% of all cancer deaths, occupational hazards such as exposure to silica dust, asbestos, radon and air pollution are also risk factors [85]. Five randomized cancer screening trials were initially conducted using chest radiography as primary imaging modality [86–90]. However, none of them resulted in a substantial reduction of pulmonary cancer mortality rate. Chest radiography can detect nodules in the advanced stages, but it is not sensitive enough to detect nodules as small as 1 mm [91, 92]. With the advent of low dose multi-slice computed tomography (LDCT), detection of small pulmonary nodules is possible in much earlier stages due to its high sensitivity and volumetric characterization.

Using LDCT, the national lung cancer screening trial (NLST) reported a substantial reduction of 20% in the lung cancer mortality rate. The NLST is the largest screening trial in the world, including 53 454 participants from 33 centers in the United States [93]. Inspired by the substantial outcome of NLST, the European Society of Radiology (ESR) and the European Respiratory Society (ERS) have also provided new recommendations for lung cancer screening in Europe [94]. Since then, several randomized screening trials have been initiated across the world, including the Dutch-Belgian screening trial (NELSON) with 15 822 participants, the largest study in Europe and the second largest in the world [95]. An overview of the randomized controlled trials conducted for lung cancer screening in Europe and the United States is presented in [96].

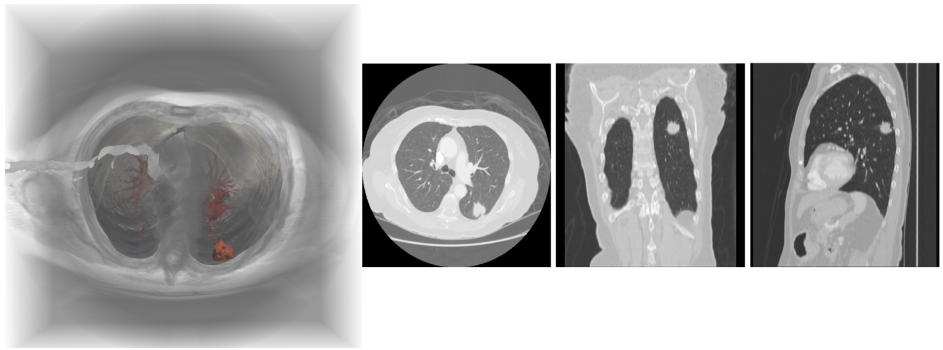


Figure 4.1 A CT volume with its three orthogonal planes: axial (left), coronal (middle), and sagittal (right) is showing an example of a large juxta-pleural nodule.

Computed Tomography

CT is commonly regarded as the primary imaging modality for the diagnosis of thoracic impairments. Its inherent high-contrast resolution allows distinguishing tissues that differ in physical density by less than 1% in comparison to 10% in conventional radiography [97]. The CT scanners incorporate a radiation source and a set of detectors. The radiation source along with the detector rotate around the patient's body for measuring the attenuation of the radiation through the body at different angles. It uses X-rays to generate 2D cross-sectional slices of the body. During acquisition, a thin axial section of a patient is imaged

by transmitting X-rays through this section from different directions. In such a way, several continuous cross-sectional slices are acquired to reconstruct a three-dimensional image using a filtered back-projection technique. The smallest 2D element in a slice corresponds to a pixel whereas the smallest 3D element in a volume corresponds to a voxel. The size of a voxel can be isotropic (uniform in all three dimensions) or anisotropic (non-uniform). The CT scans are often of anisotropic resolution i.e., voxels where $\Delta z > \Delta x \wedge \Delta x = \Delta y$.

All volumetric scans in this chapter are from the thorax region. One example of a CT volume with three orthogonal planes, i.e., axial, coronal, and sagittal are shown in Fig. 4.1. These scans typically have the dimension of $512 \times 512 \times N_z$ voxels, where N_z=100 – 500 is the total number of 2D slices in the volume. The CT scanners usually create images with an in-plane pixel resolution between 0.4 – 0.9 (in the *x*-, *y*- directions) and a resolution 0.6 – 5 mm (*z*- direction). The intensity of each voxel in a CT scan corresponds to the attenuation coefficients density of the tissue type. These attenuations are transformed on the Hounsfield scale. The Hounsfield scale ranges from - 1000 to + 4000 Hounsfield Units (HU), where each value corresponds to the beam attenuation of different tissues. It is likely that a voxel in a CT scan with thick slices might cover multiple tissues, resulting in an averaged HU value of the contained tissues for the voxel. This artifact is typically referred to as partial volume artifacts and responsible for the blurred boundaries between tissue regions [98]. Also, CT imaging tends to suffer from intensity noise, star or streak artifacts caused by metallic implants, motion blur caused by patient movements, and equipment malfunctions [98].

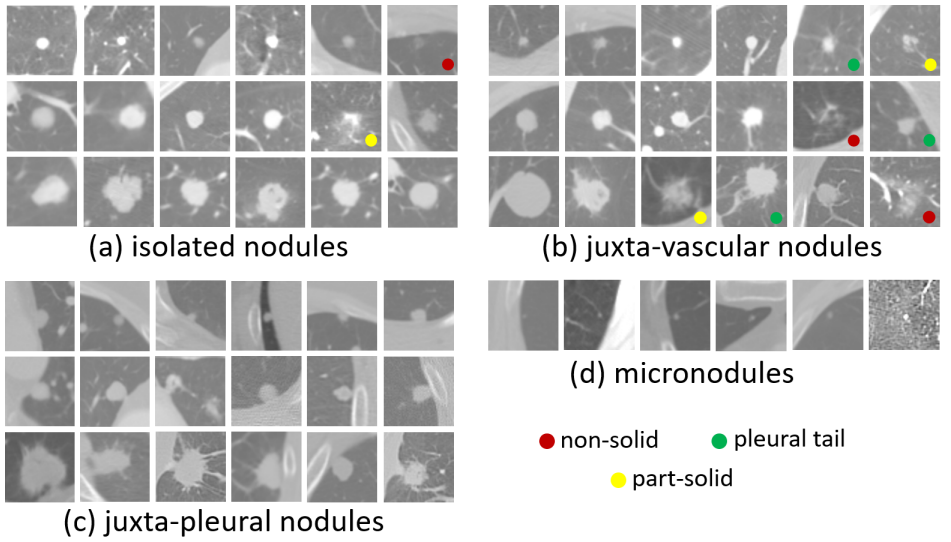


Figure 4.2 Different types of pulmonary nodules. The first, second, and third rows of a, b, and c are respectively the nodule candidates of small-size, medium-size, and large-size.

Pulmonary Nodules

Pulmonary nodules can be benign (non-communicable extensive mass of tissues) or malignant (extensive mass of tissues that can spread to the other body parts) nodules. Pulmonary nodules are radiologically characterized as the *well-* or *poorly-* defined round or oval opacities with diameters up to 30 mm [99]. The nodules on CT scans manifest high variability regarding tissue appearances, size, and shape. Some examples of different types of pulmonary nodules on CT scans are shown in Fig. 4.2.

Based on the tissue appearance, nodules can be radiologically delineated as solid, non-solid, and part-solid nodules [99]. Solid nodules exhibit a homogenous soft-tissue attenuation on the CT scans; non-solid nodules are manifested as irregular cavities with a hazy attenuation, and part-solid nodules depict the characteristics of both solid and non-solid nodules. Although part- or non-solid nodules are less prevalent, they articulate a high likelihood of malignancy as compared to the solid nodules [100]. Based on the shape, nodules can be characterized as well-circumscribed (or isolated), juxta-vascular (vessels), pleural (lung lobes border) tail, and juxta-pleural. The well-circumscribed nodules are delineated as the circular mass of tissue without any connections to vasculature, whereas vascularized nodules are explicitly connected to the surrounding vessels. The pleural tail nodules are connected both to pleura and a thin structure (vessel), whereas the periphery of the juxta-pleural nodule is connected to the pleural surface.

In addition to these radiological hallmarks, nodules are also delineated on the basis of their sizes as micronodules, small, medium, and large nodules. Micronodules are characterized as well-defined solid nodules < 3 mm and are associated with silicosis [99, 101]. Silicosis is one of the common and incurable occupational abnormalities following long and continuous exposure to silica dust. Progression of these lesions could lead to lung cancer [101]; thus, early detection is an inevitable requisite. Small nodules range from 3 mm to 6 mm, medium nodules range from 6 mm to 10 mm and are of radiological interest for the early manifestation of lung cancer. Large nodules > 10 mm pose a higher likelihood of malignancy and are often considered for further clinical diagnosis [102, 103].

Consolidation of CAD to Detect Pulmonary Nodules

The ability of CT to characterize structures > 1 mm makes it as an absolute choice over chest radiography for the early manifestation of thoracic impairments such as silicosis and lung cancer. Although the sub-millimeter resolution of helical CT scans certainly facilitates the radiological diagnostic procedure, manual interpretation of a large number of images for nodule detection is still labor-intensive and could take up to 10–15 minutes/scan (Publication C). In addition, high similarities of nodules with surrounding anatomical structures, e.g., cross-sectional vessels, further complicate the manual assessment, and contribute to the performance variability among radiologists in the detection

of nodules [104–106]. Increasing amounts of imaging data from the ongoing and future screening trials show an unequivocal requisite of CAD to assist radiologists. Acknowledging the exigence of CAD, the ESR and the ERS have explicitly recommended the adoption of the CAD for lung nodule evaluation [94].

The CAD frameworks can be subsumed either as the first, second or concurrent reader [107]. As a first reader, the CAD provides the clinicians with plausible candidates to perform further analysis on them only. This setting potentially results in less reading efforts for clinicians; however, this is precarious since it might overlook some of the suspicious abnormalities. In a second reader scenario, the clinicians perform the usual diagnostic procedure to identify suspicious abnormalities first and then use the CAD findings in a subsequent step for reconsideration of the highlighted markings. Although this setting certainly improves the detection performance, it is still ineffective since it will be time-consuming and labor-intensive in routine practices and even more so in large-scale screening trials. In the concurrent reading, the clinicians perform the detection procedure alongside the CAD frameworks simultaneously. Such a setting allows clinicians to accept or reject the CAD findings and to combine their markings with the CAD markings simultaneously.

In 1998, the US Food and Drug Administration (FDA) approved the integration of the first CAD framework as the second reader for breast cancer detection in mammograms [108]. Acknowledging that the second reader scenario is still laborious, researchers have lately proposed some rigorous methods and protocols to envisage CAD as a concurrent reader as well as a first reader [106, 109–114].

For the potential realization of the CAD systems as a first reader in the cancer-screening trials, it is imperative that the CAD should exhibit a high sensitivity with a low false positive rate (FPR) for all sizes of nodules. It should also comprehend the dissimilar image acquisition parameters and annotation criteria of different screening sites.

4.2. CAD for the Early Manifestation of Lung Cancer

During the last decade, a substantial amount of research has been put into CAD to detect different sizes of nodules [115–120]. By exploiting the discriminative capabilities of DNN as an FP reduction module, CAD researchers have shown promising results for the detection of pulmonary nodules [18, 121–123]. A summary of some existing methods is listed in Publication C. Although existing CAD systems have shown considerable detection performance; they are still limited to exhibit a high sensitivity at low FP rate and also often miss detecting subgroups of suspicious nodules. This is because most of the methods are either only tested on small datasets or limited to a single dataset, consequently limiting their potential inference as an assistive tool. Aiming at alleviating the limitations associated with the current lung nodule CAD systems, an automated workflow is developed to reduce the performance gap of CAD systems for the detection of different sizes of nodules. The proposed CAD with an efficient MLP based

FP reduction module is tested on four heterogeneous datasets consisting of CT scans from both cancer-screening trials and clinical routine (Publication C).

Method

The developed CAD system is comprised of two phases (or modules): 1) primary module and 2) final module, as shown in Fig. 4.3. A comprehensive overview of the developed CAD can be found in Publication C and is briefly summarized here. The primary module (candidate-screening stage) aims at locating the plausible nodule candidates in the segmented lung regions whereas the final module (FP reduction stage) aims at reducing the FPs using discriminative features within an optimized NN-based classifier.

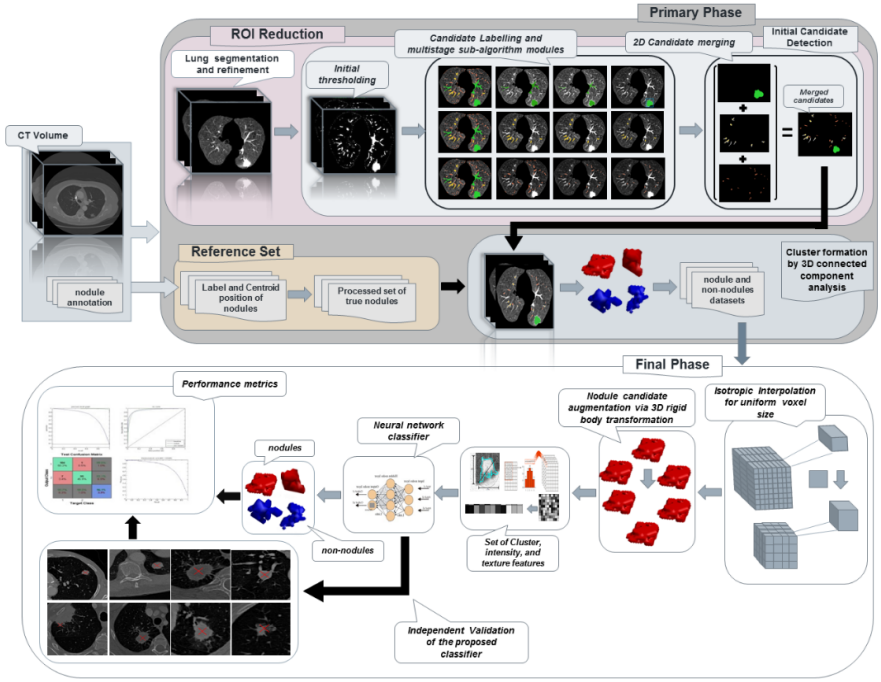


Figure 4.3 Overview of the developed CAD pipeline. The grey and black arrows show the general flow and parts for performance evaluation of the CAD pipeline, respectively.

In the primary module, the lungs are first extracted as the region of interest (ROI) using a threshold-based method. The method consists of the steps: *lung masks extraction*, *thorax region mask extraction*, *thorax region removal*, *trachea removal*, *left and right lungs separation*, *segmentation refinement*, and *grayscale masks extraction*. Although the intensity-based threshold methods are computationally less expensive and well-suited for isolated nodules, they often fail to include the juxta-pleura nodules in the segmented lungs due to their similar intensity characteristics to the pleural surface. The pleural surface of the segmented lungs often appears as holes when the juxta-pleura nodules

are excluded, and thereby potentially entails for a further lung segmentation refinement step. However, the refinement step globally enlarges the area of segmented lungs and could include irrelevant anatomical regions (such as hilar and lung borders) in the segmentation; yet, it still ensures the potential inclusion of pleura nodule candidates.

The next step in the primary module is to localize the plausible nodule candidates within the extracted lung regions. This is a challenging task since nodules pose a high variability regarding shape, size, density, contextual surrounding, and orientation. Considering that the higher threshold level often attenuates the small juxta-pleura nodules and that the lower threshold introduces a large number of FPs, an intensity threshold of -700 HU is empirically determined and applied to identify the initial candidate regions. Thereafter, three sub-algorithm modules are implemented for the detection of *small candidates* ($3\,\text{mm} \leq \text{diameter} < 6\,\text{mm}$), *medium candidates* ($6\,\text{mm} \leq \text{diameter} < 10\,\text{mm}$), and *large candidates* ($\text{diameter} \geq 10\,\text{mm}$), respectively. Each sub-algorithm module comprises multiple steps of feature-based thresholding and morphological operations. The sub-algorithm modules for the detection of both small- and medium- candidates proceed in three steps, whereas the module for the detection of large candidates consists of six stages. Since the morphological characteristic of large nodules is comparatively distinctive from their anatomical surroundings, a combination of morphological opening and feature-based thresholding should be enough to detect most of the large nodules. However, the large nodules are often interweaved with vessels and pleura structures, and thereby potentially complicates the parameter selection process.

The candidates detected in the initial module follow an imbalanced data distribution where the nodule dataset is much smaller than the non-nodule dataset. Acknowledging that this imbalance can negatively influence the training of the classifier, data augmentation techniques are applied only on the existing nodule dataset to balance the distribution. The nodules are composed of solid tissue and should not be extensively skewed, and therefore only translational and rotational transformations are performed on each nodule candidate. By doing so, the classifier is modeled to learn *translational*-, and *rotational*- invariant features.

In the final module, feature extraction is performed to discriminate between the nodules and non-nodules candidates. The computed features are associated with the intensity using raw pixel values, cluster (or morphology), and texture characteristics of the candidates and account for altogether 515 features. Due to variability in the sizes of the candidates, it is challenging to compute a feature vector of the same size for each candidate. An upgraded voxel-based approach is developed to quantify the density characteristic of a candidate. In this approach, a candidate is first isotropically resampled using cubic interpolation and then further resized to an object $I_{10\times10\times5}$. Although such transformations could influence the density characteristic of the candidates, the intrinsic reconstruction mechanism of the neural networks potentially enables them to handle the unusual averaging effects to some extent. Seven morphological features are computed

to determine the morphological characteristic of candidates. The morphological feature set consists of diameter, volume, solidity, eccentricity, and size of the segmented candidate in the x-, y-, and z- dimensions. Before computing the texture features, the densities of the isotropic candidates are quantized to a smaller number of gray levels using the Lloyd-Max quantization algorithm. After that, eight gray level co-occurrence matrix (GLCM) features (energy, contrast, entropy, homogeneity, correlation, sum average, variance, and dissimilarity) are calculated for the quantized candidate.

Once the features are computed, the candidates are classified using an MLP classifier. All features are normalized to zero mean and unit standard deviation. Before training of the classifier, the dataset is randomly split into a training set to learn the parameters, a validation set to optimize the training parameters, and a test set to determine the performance of the optimized parameters. The augmentation scheme is applied separately for each set to ensure the independence of the training set from the validation and test sets. The weights of the network are optimized using a scaled conjugate gradient descent algorithm in 3000 iterations. Softmax loss function (cross-entropy error loss) is used to measure the loss. L2 regularization and early stopping criteria are applied to control the potential overfitting and to improve the network generalization on the test set.

Material and Evaluation Criteria

The developed CAD system is validated using CT scans (DICOM formatted) from the four publically available datasets Lung Image Database Consortium/Image Database Resource Initiative (LIDC/IDRI), ELCAP, PCF, and SPIE-AAPM. The datasets are highly heterogeneous regarding image acquisition and reconstruction parameters. The LIDC/IDRI is one of the most referenced and largest publically available dataset, consisting of 1018 CT scans [124]. The CT scans with slice thickness > 3 mm are rejected due to their inadequate quality in the clinical and screening trials, resulting in 899 CT scans. In a two-phase annotation process, four radiologists delineated nodules < 3mm, non-nodules, and nodules ≥ 3 mm in every CT scan. The nodules that are accepted by at least three radiologists are selected, resulting in a set of 1 390 nodules in the reference set. The diameter, volume, and nodule boundaries of each nodule candidate are averaged to deal with the variation in the annotation of multiple readers. The ELCAP [125] dataset consist of 50 LDCT scans with 403 nodules annotated by two radiologists. The PCF [125] dataset consists of 33 CT scans (from three different subsets) with 40 nodules annotated by two radiologists. The SPIE-AAPM [126] dataset consists of 70 CT scans with 83 nodules annotated by two radiologists.

The candidate-screening stage is evaluated using a criterion that if a candidate hits within a three pixels range of the center-of-mass of the respective nodule in the reference set, it is considered as a TP candidate, else an FP candidate. The performance of the FP reduction stage is evaluated using the area under the ROC curve (AUC_R) and the competition performance metric (CPM). The CPM [127]

determines the average sensitivity of the CAD system at different operating points of the FROC that are commonly used in the screening trials points.

Results and Discussion

The CAD system is designed to detect a broad spectrum of nodule candidates. The overall performance of the CAD system is influenced by the performance of the primary module. The optimization of the lung segmentation is an important step to maximize the inclusion of nodules in the primary module. Including a morphological based refinement step, the detection rate of the candidate-screening step is improved. Examples of the successful lung segmentation refinement are shown in Fig. 4.4(a). Although the refinement step partially includes the nodules in some cases, such lesions are still detected by the candidate-screening step.

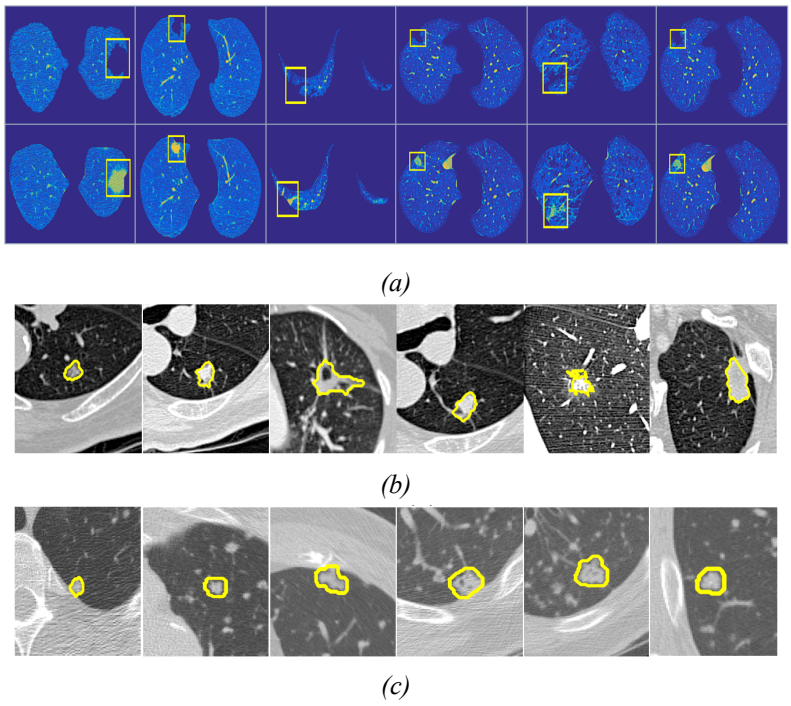


Figure 4.4 Examples of (a) lung segmentation refinement method. The first row shows the CT slices after the initial lung segmentation method wherein the pleural nodules are not included. the second row shows the corresponding CT slices after the refinement process. (b) Examples of different types of nodules detected by CAD system in SPIE-AAPM dataset. (c) Nodules not marked in the ground-truth list of PCF subset but detected by the proposed CAD.

The candidate-screening step is initially optimized using 15% of randomly selected CT scans from the LIDC/IDRI dataset and then validated on the full dataset along with the three other datasets. Table 4.1 shows that the developed method can locate most of the different sizes of nodules in the datasets. Such

high detection rate is realized at the cost of a large number of FPs. Some examples of the detected nodules in the SPIE-AAPM dataset are shown in Fig. 4.4(b). Figure 4.4(c) shows examples of unannotated nodules from the PCF dataset that are still detected by the CAD system.

Table 4.1 Performance of initial candidate detection module on different dataset. The results for independent dataset marked with an asterisk are evaluated from FROC curve.

Candidate set	Scan	True nodules	Detected nodules	Sensitivity	FPs
LIDC−IDRI					
Small candidate	899	603	569	94.3%	502 957
Medium candidate		498	476	95.6%	176 856
Large candidate		289	273	94.4%	51 673
Combined set		1,390	1,318	94.8%	731 486
Independent dataset*					
ELCAP	40	396	327	82.6%	27 239
PCF	33	40	35	87.5%	18 058
SPIE-AAPM	70	83	69	83.2%	50 462

On the 899 CT scans from the full LIDC/IDRI dataset, the developed CAD system achieved an overall sensitivity of 85.6% at 8 FPs/scan and an AUC_R of 0.957. On the given 153 CT scans from the other three datasets, the CAD system achieved an average sensitivity of 68.4% at 8 FPs/scan. The performance of the CAD system on the three datasets is adequate, especially considering that not a single of these CT scans is used for the optimization of the CAD system, so the test set is really representative for generalization to a real clinical situation. Sensitivities at seven operating points along with average CPM and AUC_R for each dataset are summarized in Table 4.2.

Table 4.2 Quantitative performance of CAD system on different CT scan datasets. Sensitivities at 7 operative points and area under ROC are also listed.

FPs/scan	1/8	1/4	1/2	1	2	4	8	Avg.	AUC_R
LIDC/IDRI	0.531	0.629	0.790	0.835	0.843	0.848	0.856	0.763	0.957
SPIE	0.194	0.305	0.442	0.628	0.640	0.640	0.663	0.501	0.831
ELCAP	0.313	0.538	0.566	0.629	0.712	0.712	0.718	0.598	0.879
PCF	0.225	0.330	0.387	0.461	0.540	0.596	0.689	0.463	0.804

4.3. CAD for the Early Manifestation of Silicosis

In contrast to CAD workflow for the early manifestation of lung cancer is an active field of research, automated detection of micronodules in thoracic CT scans still an underexplored domain. Their high similarities to cross-sectional

vessels and obscure symptoms, impose challenges in the detection task. While acknowledging the performances of CNN's in the early manifestation of lung cancer, leveraging the discriminative power of CNN's to deal with the challenges associated with micronodules is a promising choice.

In order to facilitate the interpretation of thoracic CT scans for silicosis detection, a CAD system employing a 3D CNN as a FP reduction module is developed for the detection of micronodules (Publication D).

Method

The outline of the developed CAD system is shown in Fig. 4.5. It follows a two-stage classification process where the potential candidates are initially localized by the candidate-screening module and are further scrutinized by the FP reduction module. The comprehensive overview of the developed CAD is elucidated in Publication D.

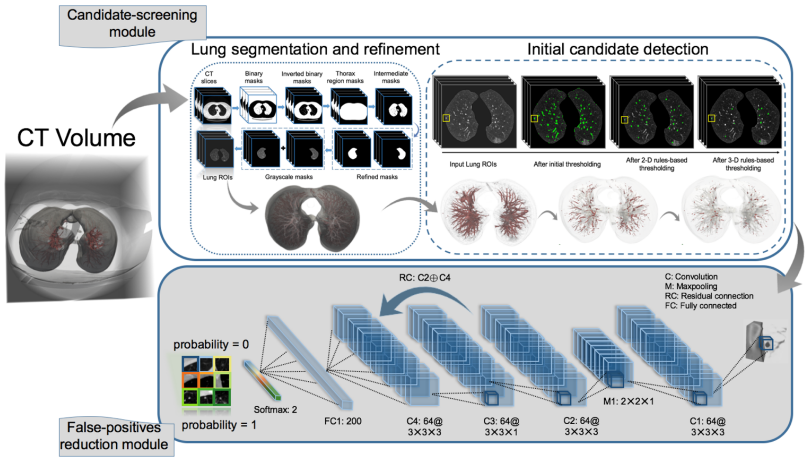


Figure 4.5 An overview of the proposed CAD system. The system is divided into a candidate-screening and a false positive reduction module. Initial candidates are detected from the segmented lung ROI's using 2D and 3D features-based thresholding operations. The false positive reduction module is implemented using a 3D CNN. The architecture is shown using an example of an extracted 20×20×7 voxels candidate.

In the candidate-screening module, the CT scans are resampled to an isotropic voxel size of 0.6 mm^3 using cubic interpolation. Isotropic resolution is often beneficial for the implementation of generic image analysis operations while dealing with the inconsistent slice-thickness across different CT scans. Next, the lung regions are extracted using the method described in Section 4.2. Once the lung regions are obtained, the plausible micronodule candidates are localized. This is challenging due to the high similarity of micronodules with cross-sectional vessels in CT scans. Although a sliding window method can be employed where each voxel is a potential center, it is impractical since it yields few positives and a large amount of FP. Instead of relying on generic methods such as selective

search to reduce the search space, an algorithm is specifically designed to identify plausible micronodule candidates. First, an intensity threshold of -700 HU is applied to identify the initial candidate regions. Next, two 2D shape descriptors (area and eccentricity) and two 3D shape descriptors (elongation and sphericity) are used to discard FP candidates.

In the FP reduction module, the original anisotropic CT scans are used so that this module can also be employed independently of the former module. Small nodules (up to 4 mm) in anisotropic CT scan typically range over up to nine voxels and four slices due to the lower resolution in the z-direction. Considering that the CT scans are often sampled more densely in the x-, y- directions than in the z-direction, a bounding box of $20\times20\times7$ voxels is extracted to include sufficient contextual information as an input to the classifier. To deal with the class imbalance, the dataset of true candidates is augmented using translation, flipping, and rotation transformations. Before the training, the intensities of each candidate are clipped to the interval (-1000, 1000 HU) and normalized to the range (0, 1). After that, the developed 3D CNN exploits the contextual information of the candidates to yield the final prediction.

To encode the discriminative representations (feature maps), the 3D CNN is developed by cascading four convolutional blocks, one pooling layer, one fully connected (dense) layer, and one Softmax layer. The convolutional block consists of one 3D convolutional layer, one BN layer [77], and one ReLU [56] nonlinear activation. Each convolution block generates 64 feature maps by convolving $3\times3\times3$ filters, except the third convolution which convolves $3\times3\times1$ filters to generate 64 feature maps. The maximum pooling layer of size $2\times2\times1$ downsamples the output of the first convolution block. The fully connected layer consists of 200 neurons and is followed by a Softmax layer to predict the final probability distribution of the input candidate. To further improve the performance of the classifier, residual (skip) connection (discussed in Chapter 2) is employed by adding the output of the second convolution block to the input of the fourth convolution block.

The weights of the 3D CNN are initialized using Glorot normal distribution [23], and the biases are initialized with zeros. The weights are adaptively updated in mini-batches of 128 candidates using the RMSProp optimizer [32]. The network is trained for 50 epochs in a five-fold cross-validation scheme with a learning rate of 0.01. A dropout [57] layer with a probability of 0.5 is implemented on the output of the dense layer to avoid the overfitting. The error loss is analyzed using Softmax loss function. The 3D CNN is implemented using Tensorflow backend in Keras.

Results and Discussion

The developed CAD system is validated on 598 CT scans from the LIDC/IDRI dataset, including 872 micronodules annotated by at least two radiologists as discussed in Section 4.2. The overall performance of the developed CAD system is evaluated using the criteria discussed in Section 4.2. Before that, multiple

experiments are conducted to analyze the impact of three crucial parameters on the performance of the 3D CNN configuration. The three parameters are associated with the effect of 1) the BN layer, 2) residual connection, and 3) size of the receptive field across *z*- direction.

The effect of the BN layer is evaluated by repeating the experiments without including it in the 3D CNN with a residual connection. It is observed that the inclusion of the BN layer improves the training performance of the model and that converges faster in comparison to the one without it (Publication D). Also, the 3D CNN with BN layer achieves a higher validation accuracy due to its regularizing effect and stable propagation of gradients. The impact of the residual connection is evaluated by repeating the experiments without including it in the 3D CNN model. The inclusion of residual connections exhibits high discriminability due to that it enables gradients to flow smoothly through the skip connection during the backward pass of the training. The effect of the receptive field across *z*- direction is evaluated by conducting experiments using an elongated volume of $20\times20\times20$ for a new 3D CNN model. The architectural design of this model is described in Paper IV. The size of the receptive field potentially influences the final predicted probability of the classifier. The large receptive field can possess more redundancy due to relatively larger contextual surrounding, and thereby more generalized to ambiguous contextual information.

Table 4.3 Quantitative performance of CAD system with different CNN settings on the LIDC/IDRI dataset. Sensitivities at 7 operative points, CPM and AUC are also listed.

FPs/scan	1/8	1/4	1/2	1	2	4	8	CPM	AUC_R
3D CNN_r	0.549	0.629	0.680	0.743	0.792	0.832	0.867	0.727	0.988
3D CNN	0.505	0.571	0.642	0.707	0.772	0.793	0.845	0.691	0.957
3D CNN_e	0.280	0.401	0.485	0.593	0.682	0.760	0.814	0.573	0.943
Shallow NN	0.125	0.330	0.377	0.441	0.510	0.556	0.609	0.421	0.884

Furthermore, a shallow NN trained with conventional features is compared with the CNN features. The computed set consists of 21 intensity and 6 morphological features from two spherical regions centered on the candidates. All features are normalized to zero mean using unit standard deviation. The conventional features are not able to yield a high CPM as compared to the CNN features. The traditional features are often affected by the similar intensity distribution of nearby blood vessels and tissues in the segmentation results from the candidate-screening module.

On the given 872 nodules in 598 CT scans from the LIDC/IDRI dataset, the candidate-screening module generates an average of 447 candidates/scan, including 91.6% (799/872) of all micronodule candidates. For quantitative comparison, the sensitivities at seven operating points along with average CPM and AUC_R for each CNN configuration are listed in Table 4.3. By employing the best 3D CNN configuration, the developed workflow identifies 86.7% (756/872) of micronodules at 8 FPs/scan and achieves an AUC_R of 0.988. The sensitivities

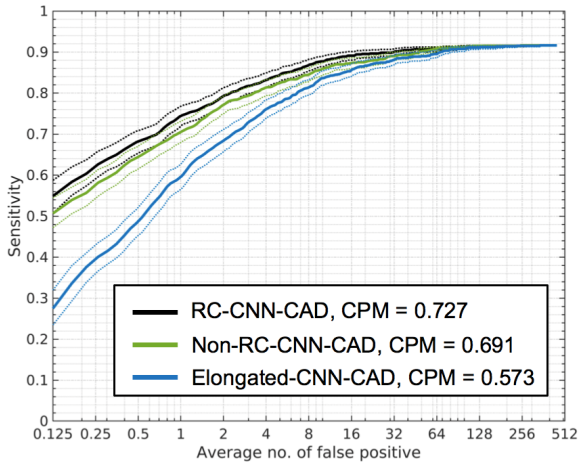


Figure 4.6 FROC curve for the different CNN configurations tested on the LIDC/DIRI dataset. The dashed curve shows the 95% bootstrap confidence interval. The number of false positives are shown on a logarithmic scale.

at seven operating points along with average CPM and AUC_R for each CNN configuration are listed in Table 4.3. Most of the FPs detected at 1 FP/scan are the small vessels, nodule-like structures, and scaring. All these structures have the same characteristics as micronodules. Some examples of detected micronodules and FPs are shown in Fig. 4.7.

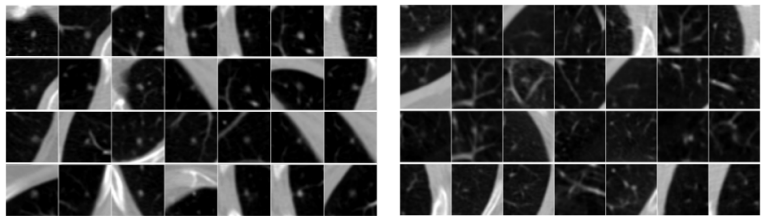


Figure 4.7 Examples of lesions detected by the CAD system. The left and right set of lesions are respectively the micronodules and the FP candidates detected at 1 FP/scan.

5. CLASSIFICATION OF VASCULAR SKELETON CROSS-SECTIONS IN CTA IMAGES

Given the complexity and abundance of CT angiography (CTA) imaging data, radiologists continually seek faster and more accurate computerized methods for segmenting vascular structures. The existing methods are mostly skeleton-based and require a vessel skeleton extracted prior the segmentation. The previous method developed by the collaborators employs hand-engineered filters for fast vascular skeleton extraction. This method generates enormous amounts of FPs and also operates multiple times to detect different sizes of vascular nodes.

To simplifying the existing workflow, a patch-based 2D CNN classifier is developed that classifies cross-sections of different sizes of vascular nodes in a single pass (Paper V). Instead of developing a workflow from scratch, this work focuses on improving the node-classification step and is consolidated with already existing workflow. This chapter summarizes the background, material, methods, results, and contribution presented in the appended publication E.

5.1. Overview

Vascular image analysis is essential both for diagnostic and meticulous treatment planning [128]. Modern non-invasive vascular imaging techniques such as CTA provide enhanced volumetric characterization, which allows clinicians to perform a diagnosis at the high-resolution level and thereby facilitating the diagnostic procedure. Also, non-invasive surgeries benefit patients by reducing the risk of complications and improving their comfort [129]. However, such benefits are realized at the cost of large amounts of imaging data. Acknowledging that manual interpretation of such an amount of data is monotonous, labor-intensive and error-prone, automation is highly desirable for fast vascular skeleton extractions and accurate vessel segmentation in the CTA.

Such an automated workflow for fast vascular skeleton extraction has previously described [130]. Tubular vessels are usually visualized as bright elliptical-like regions on darker background of 2D orthogonal CT slices. Due to the injected contrast medium, the intensity of vessels is higher than the surrounding tissue intensities, however, is equal to the intensity of spongy trabecular bone. One step in the previous workflow employs a set of

hand-engineered filters for classifying 2D orthogonal cross-sections of vascular and non-vascular nodes. The hand-engineered filters achieve a satisfactory detection performance but also introduce an enormous amount of FPs. Moreover, this method operates in multiple iterations with a different set of filers since the same set of filters were not suitable for all types and sizes of the vessels. Given the adequate performance of the CNN's in reducing FPs, a patch-based 2D CNN classifier is employed to classify different sizes of vascular cross-sections of healthy as well as diseased vessels. In the existing workflow, the hand-engineered filters are replaced with the developed classifier, which significantly reduces the FPs in a single algorithm pass.

5.2. Method

The workflow is composed of four steps as shown in Fig. 5.1. This chapter solely focuses on the *node-candidate classification* step whereas the remaining three steps [129] are briefly described in Publication E.

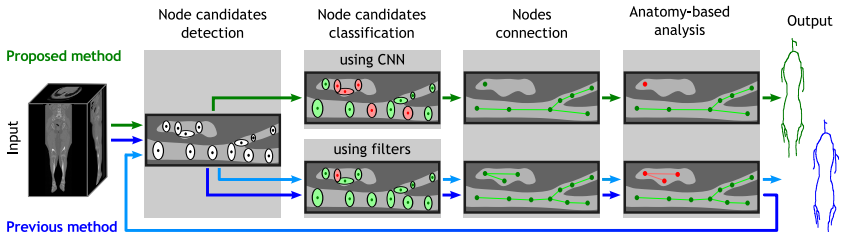


Figure 5.1 The pipeline of the proposed method (green, top) produces the final vascular skeleton in one algorithm pass compared to the pipeline of the previous method [130] (blue, bottom) which detect skeletons of larger arteries in the first iteration and adds the skeletons of smaller arteries in the second iteration

Initially, patches of 31×31 pixels centered on the node candidates are extracted as the input to the classifier. The pixel values of the patches are kept in HU not to lose the density characteristics of the nodes. Given the heterogeneity of vascular cross-sections regarding contextual surrounding, shape, size, and orientation, it is worthwhile to model such variability in a classifier for additional performance gain. However, the dataset of true vascular nodes is much smaller compared to the dataset of non-vascular (or false vascular) nodes, and thereby negatively influences the performance. To overcome this issue, new samples are generated using the translation of 1 pixel in both x- and y- directions, horizontal and vertical flipping, and six random rotations (0-180°) transformations, resulting in 10 augmented variations for each vascular node. Acknowledging that the initially detected nodes are often decentered from local anatomical structures, such transformations capacitate the CNN with orientation-independent features. Once the data is prepared, the candidates are classified using a 2D CNN classifier.

Before the training, the patches are normalized by subtracting the mean and dividing by the standard deviation.

The CNN classifier consists of four convolutional layers and two max-pooling layers as shown in Fig. 5.2. Firstly, the input patches are padded with a two-pixel thick frame of zeros to keep the spatial sizes the same as the original size after the first convolutional layer. The first and second convolutional layers generate 32 feature maps respectively using 3×3 convolutions. The third and fourth convolutional layers generate 64 feature maps respectively using 3×3 convolutions. Each convolution layer is followed by a BN [77] layer. The max-pooling layer downsamples the feature maps by selecting the maximum feature response in windows of size 2×2. The fully connected (dense) last layer consists of 512 neurons and is followed by a softmax layer to predict the final probability distribution of the input candidate. ReLU [56] nonlinear activation is applied after every convolutional and dense layer.

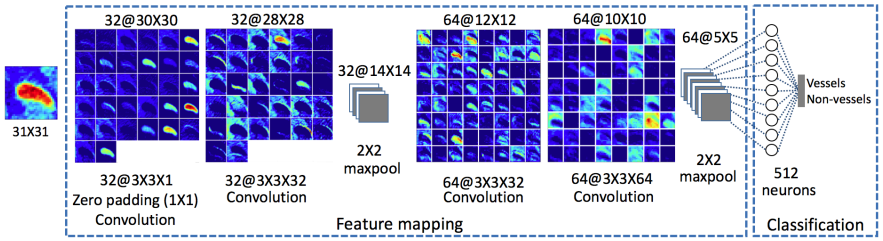


Figure 5.2 A schematic overview of proposed CNN classifier, showing the output of each convolution filter applied to an example patch of a vessel. Here, the grayscale intensities are shown in color for suitable visualization.

The weights of the CNN are initialized using Glorot normal distribution [23] and the biases are initialized with zeros. The weights are adaptively updated in mini-batches of 128 candidates using the RMSProp [32] optimizer. The training continues for 20 epochs in a five-fold cross-validation scheme with a learning rate of 0.01. A dropout [57] layer with a probability of 0.25 is implemented on the output of each pooling layer whereas a probability of 0.5 is used on the output of the dense layer. The error loss is measured using the softmax loss function. The 2D CNN is implemented using Theano backend in Keras.

Material and Evaluation Criteria

The classifier is validated on 25 CTA volumes of the lower limbs. The dataset was obtained from the clinical routine of the Radiology department at University Hospital, Linköping, Sweden. An expert radiologist (author ÖS in Paper V) used a semi-automatic segmentation tool to delineate the ground-truth in four CTA volumes. In these four annotated volumes, the node-candidate detection step identifies 352 523 different sizes of cross-sections of vascular and non-vascular nodes. These cross-sections are categorized into small cross-sections (<4 mm), medium cross-sections (4-20 mm), and large cross-sections (>20 mm).

A reference set is constructed using the medium-sized cross-sections only, which consists of altogether 138 302 cross-sectional patches (24 625 vascular and 113 677 non-vascular). The smaller cross-sections are excluded due to their insufficient contextual information during training which may influence generalization of the network. The larger cross-sections are also excluded due to their larger size than the chosen patch size. However, both, smaller and larger candidate cross-sections are used for testing the CNN classifier. Given the limited amount of annotated samples, the reference set is divided into two subsets: *model-development subset* and *model-evaluation subset*.

The model-development subset, consisting of 20 000 candidates from each class, is divided into training, validation, and test sets. The training and validation sets are used for the cross-validation, and the test set is used for the model selection. The training set consists of 12 000 candidates from each class, whereas both the validation and test sets consist of 4 000 candidates from each class.

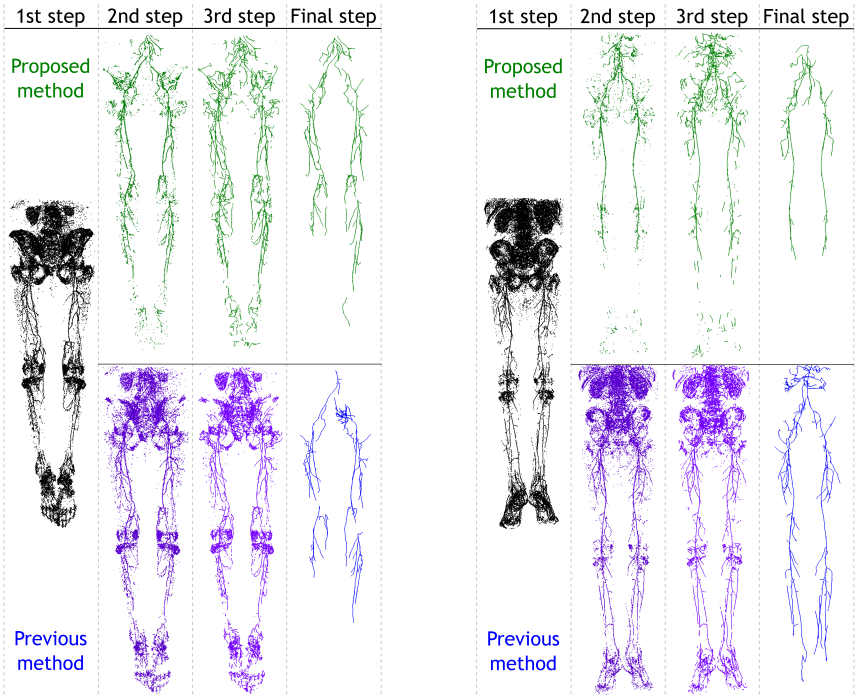
The *model-evaluation subset*, consisting of 4 625 vessels and 93 677 non-vessels samples, is used for the quantitative evaluation of the workflow employing the CNN classifier. The reason for keeping the distribution of samples in the model-evaluation subset imbalance is to test the workflow in a real clinical scenario. The performance is quantified regarding precision, recall, AUC_{PR}, and F-score. The remaining unannotated twenty-one CTA volumes are used for the qualitative evaluation and comparison purposes.

Table 5.1 Comparative evaluation of CNN classifier and hand-engineered filters.

	Set	**Prec.**	**Rec.**	**F-score**	AUC_{PR}
CNN	small	0.66	0.71	0.65	0.69
	medium	0.81	0.83	0.82	0.90
	large	0.70	0.75	0.72	0.86
filters	small	0.29	0.78	0.42	–
	medium	0.28	0.89	0.43	–
	large	0.06	0.58	0.11	–

5.3. Results and Discussion

For the 98,302 medium-sized vessel nodes in the model-evaluation subset, the workflow employing the CNN classifier yields a precision of 0.81 compared to 0.28 for the workflow utilizing hand-engineered filters. In addition, the developed CNN classifier achieves an AUC_{PR} of 0.90 for the medium-sized nodes. Given that the CNN classifier is trained using only one candidate group, it is noteworthy that the CNN performs competently well also for the other candidate groups. Table 5.1 summarizes the performance of the classifier for each candidate group. Employing a CNN classifier is ultimately simplifying the previous vascular skeleton extraction workflow.



(a) Volume A *(b) Volume B*

Figure 5.3 Results after each algorithm step for 2 volumes; result after the 1st step is same for both methods.

For qualitative comparison, the results of both workflows for two representative volumes are shown in Fig. 5.3. The results for Volume A confirm that the suggested workflow leads to a fewer number of false-positives, including more vascular branches compared to the previous method. Comparatively, the workflow employing hand-engineered filters detects a large number of FPs, especially in the pelvic region. However, in some cases, the CNN classifier missed detecting small or diseased true vessel candidates, leading to completely missed vessel branches as shown in Volume B. This is justified by the fact that the CNN classifier needs a substantial amount of such diseased samples to classify them correctly.

SUMMARY AND CONCLUSIONS

Technologically sophisticated imaging modalities such as TEM and CT have phenomenally aided the pathologists and the radiologists in detecting and diagnosing subtle abnormalities (e.g., genetic disorders and lung cancer). However, clinicians often feel overburden by interpreting (or analyzing) images manually, and thus, they quest for CAD systems.

In this thesis, various challenges associated with CAD systems for the TEM and the CT imaging modalities have been investigated and addressed. As a result, several contributions in topics related to detection, false positive reduction (classification), and denoising have been made. Several DNN-based methods have been developed for the detection of objects in CT and CTA images to realize their potential for medical applications. Given this improved performance, the capabilities of CNN-based methods have also been leveraged for problems concerning more complex structures such as cilia in TEM images.

After introducing the current problems associated with manual diagnosis, Chapter 1 elucidated the term "CAD" as an answer to *how the diagnostic procedures can be automated*. This chapter also provided an insight into the generic structure of the CAD systems. Acknowledging the limitations of the conventional CAD systems employing hand-engineered features, this chapter also revealed the perspective of contemporary CAD systems by employing DNN as an answer to *what can be done to improve the performance further*. By explaining the technical background of DNN, the Chapter 2 discussed the different components that have been used to develop the DNN-based models in this thesis. While discussing the overall contributions of this thesis, the last three chapters answered *how the CAD systems have been developed using DNN-based methods*. In particular, this has been accomplished by focusing on the research objectives listed in Section 1.2.

Summary of Claims

This section lists the claims of novelty that were shown in this PhD work. The claims correspond to Contributions 1 – 5 and are reflected in Papers I – V.

> **Claim 1:** The proposed CNN model complements the automated workflow for the detection of cilia in low-magnification TEM images. No CNN model as an FP reduction module has been proposed or published previously for automated PCD diagnosis in TEM images.

Claim 2: The proposed novel multi-stream CNN model enhances the structural information by denoising short-exposure high-magnification TEM images. The CNN model has not been proposed previously for denoising TEM images.

Claim 3: The proposed CAD system complements the other existing CAD solutions since it can comprehend multiple sizes of nodules in CT scans without being affected by any image acquisition parameters. It is the only CAD system that has been extensively tested on four publicly available CT datasets including the largest dataset, i.e., LIDC/IDRI, and thus it is the first study conducted on such a large scale.

Claim 4: The proposed novel 3D CNN model shows promising results for the early detection of micronodules in CT scans. This is the first study to present a CAD system that employs a 3D CNN model for the detection of silicosis.

Claim 5: The proposed CNN model complements the automated workflow for vascular skeleton extraction. In comparison to multiple passes with hand-engineered filters [130], the model classifies cross-sections of different sizes of vascular nodes in a single algorithm pass. The model substantially reduces the false positives in the vascular cross-sectional images, and thus improves the overall classification performance.

Concluding Remarks and Future Opportunities

The unique capabilities of DNN have undoubtedly emerged it as the leading learning-based methods for discovering multiple levels of distributed representations. Provided this, the CAD researchers have embraced DNN in the form of CNN for several biomedical and medical image analysis tasks such as image classification, object detection, image retrieval, objects segmentation, and others. Their inherent highly discriminative capabilities have motivated the increasing research interests. In addition, CNN learns from data and often relies minimally on domain experts, and thus, making development easier and faster.

The CNN-based methods have surpassed or substantially improved the performance compared to the conventional methods for CAD, and therefore, substantiating a great potential to advance in the computerized image analysis research field. On a similar note, this thesis is an effort to facilitate the manual diagnostic procedure through computerized analysis frameworks. This thesis has elucidated the applicability of DNN-based methods for the classification and denoising of objects in TEM and CT images. It is thus evident from this thesis that DNN has competencies to penetrate multiple aspects of biomedical and medical image analysis. Given the astonishing results of CNN's on diverse applications, several directions can be drawn for future research topics.

The field of computerized image analysis has seen a transition from purely hand-engineered representations to automated discriminative neuron-crafted

representations. Currently, both paradigms are combined to deploy more sensitive CAD systems where hand-engineered features are often used for identifying plausible candidates, followed by feature extraction and their classification using CNN. Although the presented CAD system employing image-based machine learning for the detection of pulmonary nodules in CT images has shown a substantial performance, it is envisioned that the performance can be boosted further by exploiting a 3D CNN framework. On the other hand, this approach can also be taken to the next level by leveraging CNN's as a single module for both: initial candidate detection and classification, and thus, removing the requisite of feature engineering entirely.

Since research interest is transiting further to end-to-end trained CNN's for object recognition and localization simultaneously in images, it also indicates a possibility to develop CAD systems that can simultaneously locate multiple abnormalities in images. Training requires a substantial amount of labeled data to leverage the full potential of CNN's; however; it is not a case for CT-based pulmonary imaging due to the availability of sufficient data. One can efficiently use this data to train CNN's for entirely different correlated modalities such as mammography. Such research innovation is presented by [122] where they have trained a DNN model for pulmonary nodule's detection and further inferred it to detect breast cancer. However, their work is still limited to binary classification likewise the work in this thesis. This approach can be extended by exploiting a single CNN model for multi-class classification problems in pulmonary or breast imaging. Therefore, one striking conclusion can be drawn is that a single architecture is capable enough to deal with multiple tasks if trained at hand.

The accurate diagnosis of abnormalities mostly depends on both image acquisition and image analysis. Contrary to medical imaging where image acquisition has relatively improved since devices acquire images at a much faster rate and increased resolution, it is still a bit challenging with biomedical imaging modalities. The potential advantages of CNN's are not merely that they are better feature extractors. As illustrated in this thesis, CNN's can be used in other ways as well, to produce filtered images, i.e., denoising image to improve the image quality for enhanced diagnosis, and thus, illustrating their power and potential for biomedical imaging applications as well.

REFERENCES

[1] H. L. Dreyfus and S. E. Dreyfus, "Making a mind versus modelling the brain: artificial intelligence back at the branchpoint," in *Understanding the Artificial: On the future shape of artificial intelligence*. Springer, 1991, pp. 33–54.

[2] B. van Ginneken, "Fifty years of computer analysis in chest imaging: rule-based, machine learning, deep learning," *Radiological physics and technology*, vol. 10, no. 1, pp. 23–32, 2017.

[3] J.-I. Toriwaki, Y. Suenaga, T. Negoro, and T. Fukumura, "Pattern recognition of chest x-ray images," *Computer Graphics and Image Processing*, vol. 2, no. 3-4, pp. 252–271, 1973.

[4] B. Sahiner, H.-P. Chan, L. M. Hadjiiski, P. N. Cascade, E. A. Kazerooni, A. R. Chughtai, C. Poopat, T. Song, L. Frank, J. Stojanovska *et al.*, "Effect of cad on radiologists' detection of lung nodules on thoracic ct scans: analysis of an observer performance study by nodule size," *Academic radiology*, vol. 16, no. 12, pp. 1518–1530, 2009.

[5] S. Schalekamp, B. van Ginneken, E. Koedam, M. M. Snoeren, A. M. Tiehuis, R. Wittenberg, N. Karssemeijer, and C. M. Schaefer-Prokop, "Computer-aided detection improves detection of pulmonary nodules in chest radiographs beyond the support by bone-suppressed images," *Radiology*, vol. 272, no. 1, pp. 252–261, 2014.

[6] K. N. Jeon, J. M. Goo, C. H. Lee, Y. Lee, J. Y. Choo, N. K. Lee, M.-S. Shim, I. S. Lee, K. G. Kim, D. S. Gierada *et al.*, "Computer-aided nodule detection and volumetry to reduce variability between radiologists in the interpretation of lung nodules at low-dose screening CT," *Investigative radiology*, vol. 47, no. 8, p. 457, 2012.

[7] Y. Zhao, G. H. de Bock, R. Vliegenthart, R. J. van Klaveren, Y. Wang, L. Bogoni, P. A. de Jong, W. P. Mali, P. M. van Ooijen, and M. Oudkerk, "Performance of computer-aided detection of pulmonary nodules in low-dose ct: comparison with double reading by nodule volume," *European radiology*, vol. 22, no. 10, pp. 2076–2084, 2012.

[8] R. F. Brem, J. Baum, M. Lechner, S. Kaplan, S. Souders, L. G. Naul, and J. Hoffmeister, "Improvement in sensitivity of screening mammography with computer-aided detection: a multiinstitutional trial," *American Journal of Roentgenology*, vol. 181, no. 3, pp. 687–693, 2003.

[9] L. J. Warren Burhenne, S. A. Wood, C. J. D'Orsi, S. A. Feig, D. B. Kopans, K. F. O'Shaughnessy, E. A. Sickles, L. Tabar, C. J. Vyborny, and R. A. Castellino, "Potential contribution of computer-aided detection to the sensitivity of screening mammography," *Radiology*, vol. 215, no. 2, pp. 554–562, 2000.

[10] R. A. Castellino, "Computer aided detection (CAD): an overview," *Cancer Imaging*, vol. 5, no. 1, p. 17, 2005.

[11] H. Miller, "The froc curve: A representation of the observer's performance for the method of free response," *The Journal of the Acoustical Society of America*, vol. 46, no. 6B, pp. 1473–1476, 1969.

[12] D. Chakraborty, "A status report on free-response analysis," *Radiation protection dosimetry*, vol. 139, no. 1-3, pp. 20–25, 2010.

[13] K. Suzuki, "Overview of deep learning in medical imaging," *Radiological physics and technology*, vol. 10, no. 3, pp. 257–273, 2017.

[14] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. van der Laak, B. van Ginneken, and C. I. Sánchez, "A survey on deep learning in medical image analysis," *Medical image analysis*, vol. 42, pp. 60–88, 2017.

[15] K. Fukushima, S. Miyake, and T. Ito, "Neocognitron: A neural network model for a mechanism of visual pattern recognition," *IEEE Transactions on systems, man, and cybernetics*, no. 5, pp. 826–834, 1983.

[16] S.-C. Lo, S.-L. Lou, J.-S. Lin, M. T. Freedman, M. V. Chien, and S. K. Mun, "Artificial convolution neural network techniques and applications for lung nodule detection," *IEEE Transactions on Medical Imaging*, vol. 14, no. 4, pp. 711–718, 1995.

[17] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, 2012, pp. 1097–1105.

[18] B. van Ginneken, A. A. Setio, C. Jacobs, and F. Ciompi, "Off-the-shelf convolutional neural network features for pulmonary nodule detection in computed tomography scans," in *Biomedical Imaging (ISBI), 2015 IEEE 12th International Symposium on*. IEEE, 2015, pp. 286–289.

[19] W. S. McCulloch and W. Pitts, “A logical calculus of the ideas immanent in nervous activity,” *The bulletin of mathematical biophysics*, vol. 5, no. 4, pp. 115–133, 1943.

[20] D. Anderson and G. McNeill, “Artificial neural networks technology,” *Kaman Sciences Corporation*, vol. 258, no. 6, pp. 1–83, 1992.

[21] M. Minsky, S. A. Papert, and L. Bottou, *Perceptrons: An introduction to computational geometry*. MIT press, 2017.

[22] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, “Learning representations by back-propagating errors,” *Nature*, vol. 323, no. 6088, p. 533, 1986.

[23] X. Glorot and Y. Bengio, “Understanding the difficulty of training deep feedforward neural networks,” in *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics*, 2010, pp. 249–256.

[24] A. L. Maas, A. Y. Hannun, and A. Y. Ng, “Rectifier nonlinearities improve neural network acoustic models,” in *Proceedings of the IEEE international conference on machine learning*, vol. 30, no. 1, 2013, p. 3.

[25] K. He, X. Zhang, S. Ren, and J. Sun, “Delving deep into rectifiers: Surpassing human-level performance on imagenet classification,” in *Proceedings of the IEEE international conference on computer vision*, 2015, pp. 1026–1034.

[26] P. Ramachandran, B. Zoph, and Q. V. Le, “Swish: a Self-Gated Activation Function,” *arXiv preprint arXiv:1710.05941*, 2017.

[27] D. H. Hubel and T. N. Wiesel, “Receptive fields, binocular interaction and functional architecture in the cat’s visual cortex,” *The Journal of physiology*, vol. 160, no. 1, pp. 106–154, 1962.

[28] Y. LeCun, Y. Bengio, and G. Hinton, “Deep learning,” *Nature*, vol. 521, no. 7553, p. 436, 2015.

[29] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[30] N. Qian, “On the momentum term in gradient descent learning algorithms,” *Neural networks*, vol. 12, no. 1, pp. 145–151, 1999.

[31] Y. Bengio, N. Boulanger-Lewandowski, and R. Pascanu, “Advances in optimizing recurrent networks,” in *Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on*. IEEE, 2013, pp. 8624–8628.

[32] T. Tieleman and G. Hinton, "Lecture 6.5-rmsprop: Divide the gradient by a running average of its recent magnitude," *COURSERA: Neural networks for machine learning*, vol. 4, no. 2, pp. 26–31, 2012.

[33] S. Ruder, "An overview of gradient descent optimization algorithms," *arXiv preprint arXiv:1609.04747*, 2016.

[34] N. S. Keskar, D. Mudigere, J. Nocedal, M. Smelyanskiy, and P. T. P. Tang, "On large-batch training for deep learning: Generalization gap and sharp minima," *arXiv preprint arXiv:1609.04836*, 2016.

[35] M. R. Knowles, L. A. Daniels, S. D. Davis, M. A. Zariwala, and M. W. Leigh, "Primary ciliary dyskinesia. recent advances in diagnostics, genetics, and characterization of clinical disease," *American journal of respiratory and critical care medicine*, vol. 188, no. 8, pp. 913–922, 2013.

[36] A. Barbato, T. Frischer, C. Kuehni, D. Snijders, I. Azevedo, G. Baktai, L. Bartoloni, E. Eber, A. Escribano, E. Haarman *et al.*, "Primary ciliary dyskinesia: a consensus statement on diagnostic and treatment approaches in children," *European Respiratory Journal*, vol. 34, no. 6, pp. 1264–1276, 2009.

[37] R. Eliasson, B. Mossberg, P. Camner, and B. A. Afzelius, "The immotile-cilia syndrome: a congenital ciliary abnormality as an etiologic factor in chronic airway infections and male sterility," *New England Journal of Medicine*, vol. 297, no. 1, pp. 1–6, 1977.

[38] M. Boon, J. Wallmeier, L. Ma, N. T. Loges, M. Jaspers, H. Olbrich, G. W. Dougherty, J. Raidt, C. Werner, I. Amirav *et al.*, "Mcidas mutations result in a mucociliary clearance disorder with reduced generation of multiple motile cilia," *Nature communications*, vol. 5, p. 4418, 2014.

[39] J. Wallmeier, D. A. Al-Mutairi, C.-T. Chen, N. T. Loges, P. Pennekamp, T. Menchen, L. Ma, H. E. Shamseldin, H. Olbrich, G. W. Dougherty *et al.*, "Mutations in ccno result in congenital mucociliary clearance disorder with reduced generation of multiple motile cilia," *Nature genetics*, vol. 46, no. 6, p. 646, 2014.

[40] N. Rumman, C. Jackson, S. Collins, P. Goggin, J. Coles, and J. S. Lucas, "Diagnosis of primary ciliary dyskinesia: potential options for resource-limited countries," *European Respiratory Review*, vol. 26, no. 143, p. 160058, 2017.

[41] C. Werner, J. G. Onnebrink, and H. Omran, "Diagnosis and management of primary ciliary dyskinesia," *Cilia*, vol. 4, no. 1, p. 2, 2015.

[42] C. E. Kuehni and J. S. Lucas, "Toward an earlier diagnosis of primary ciliary dyskinesia. which patients should undergo detailed diagnostic testing?" *Annals of the American Thoracic Society*, vol. 13, no. 8, pp. 1239–1243, 2016.

[43] J. S. Lucas, A. Barbato, S. A. Collins, M. Goutaki, L. Behan, D. Caudri, S. Dell, E. Eber, E. Escudier, R. A. Hirst *et al.*, "European respiratory society guidelines for the diagnosis of primary ciliary dyskinesia," *European respiratory journal*, pp. ERJ–01 090, 2016.

[44] M. W. Leigh, M. A. Zariwala, and M. R. Knowles, "Primary ciliary dyskinesia: improving the diagnostic approach," *Current opinion in pediatrics*, vol. 21, no. 3, p. 320, 2009.

[45] J. Lobo, M. A. Zariwala, and P. G. Noone, "Primary ciliary dyskinesia," in *Seminars in respiratory and critical care medicine*, vol. 36, no. 2. NIH Public Access, 2015, p. 169.

[46] A. Suveer, N. Sladoje, J. Lindblad, A. Dragomir, and I.-M. Sintorn, "Automated detection of cilia in low magnification transmission electron microscopy images using template matching," in *Biomedical Imaging (ISBI), 2016 IEEE 13th International Symposium on*. IEEE, 2016, pp. 386–390.

[47] T. de Haro and P. Furness, "Current and future delivery of diagnostic electron microscopy in the uk: results of a national survey," *Journal of clinical pathology*, vol. 65, no. 4, pp. 357–361, 2012.

[48] H.-W. Ackermann, "Sad state of phage electron microscopy. please shoot the messenger," *Microorganisms*, vol. 2, no. 1, pp. 1–10, 2013.

[49] A. M. Roseman, "Particle finding in electron micrographs using a fast local correlation algorithm," *Ultramicroscopy*, vol. 94, no. 3-4, pp. 225–236, 2003.

[50] Y. Zhu, Q. Ouyang, and Y. Mao, "A deep convolutional neural network approach to single-particle recognition in cryo-electron microscopy," *BMC bioinformatics*, vol. 18, no. 1, p. 348, 2017.

[51] J. Xu, X. Luo, G. Wang, H. Gilmore, and A. Madabhushi, "A deep convolutional neural network for segmenting and classifying epithelial and stromal regions in histopathological images," *Neurocomputing*, vol. 191, pp. 214–223, 2016.

[52] 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*. Springer, 2015, pp. 234–241.

[53] T. M. Quan, D. G. Hilderbrand, and W.-K. Jeong, "Fusionnet: A deep fully residual convolutional neural network for image segmentation in connectomics," *arXiv preprint arXiv:1612.05360*, 2016.

[54] W. J. Godinez, I. Hossain, S. E. Lazic, J. W. Davies, and X. Zhang, "A multi-scale convolutional neural network for phenotyping high-content cellular images," *Bioinformatics*, vol. 33, no. 13, pp. 2010–2019, 2017.

[55] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.

[56] V. Nair and G. E. Hinton, "Rectified linear units improve restricted boltzmann machines," in *Proceedings of the International conference on machine learning (ICML)*, 2010, pp. 807–814.

[57] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *The Journal of Machine Learning Research*, vol. 15, no. 1, pp. 1929–1958, 2014.

[58] F. Agostinelli, M. R. Anderson, and H. Lee, "Adaptive multi-column deep neural networks with application to robust image denoising," in *Advances in Neural Information Processing Systems*, 2013, pp. 1493–1501.

[59] Y. Marnissi, Y. Zheng, E. Chouzenoux, and J.-C. Pesquet, "A variational bayesian approach for image restoration—application to image deblurring with poisson–gaussian noise," *IEEE Transactions on Computational Imaging*, vol. 3, no. 4, pp. 722–737, 2017.

[60] P. Milanfar, "A tour of modern image filtering: New insights and methods, both practical and theoretical," *IEEE Signal Processing Magazine*, vol. 30, no. 1, pp. 106–128, 2013.

[61] A. Buades, B. Coll, and J.-M. Morel, "A review of image denoising algorithms, with a new one," *Multiscale Modeling & Simulation*, vol. 4, no. 2, pp. 490–530, 2005.

[62] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising by sparse 3-d transform-domain collaborative filtering," *IEEE Transactions on image processing*, vol. 16, no. 8, pp. 2080–2095, 2007.

[63] P. Chatterjee and P. Milanfar, "Clustering-based denoising with locally learned dictionaries," *IEEE Transactions on Image Processing*, vol. 18, no. 7, pp. 1438–1451, 2009.

[64] Z. Wang, Y. Yang, Z. Wang, S. Chang, J. Yang, and T. S. Huang, "Learning super-resolution jointly from external and internal examples,"

IEEE Transactions on Image Processing, vol. 24, no. 11, pp. 4359–4371, 2015.

[65] L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D: nonlinear phenomena*, vol. 60, no. 1-4, pp. 259–268, 1992.

[66] J. P. Oliveira, J. M. Bioucas-Dias, and M. A. Figueiredo, "Adaptive total variation image deblurring: a majorization–minimization approach," *Signal Processing*, vol. 89, no. 9, pp. 1683–1693, 2009.

[67] F. Luisier, T. Blu, and M. Unser, "Image denoising in mixed poisson–gaussian noise," *IEEE Transactions on image processing*, vol. 20, no. 3, pp. 696–708, 2011.

[68] B. Bajić, J. Lindblad, and N. Sladoje, "Restoration of images degraded by signal-dependent noise based on energy minimization: an empirical study," *Journal of Electronic Imaging*, vol. 25, no. 4, p. 043020, 2016.

[69] K. Zhang, W. Zuo, Y. Chen, D. Meng, and L. Zhang, "Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising," *IEEE Transactions on Image Processing*, vol. 26, no. 7, pp. 3142–3155, 2017.

[70] X. Mao, C. Shen, and Y.-B. Yang, "Image restoration using very deep convolutional encoder-decoder networks with symmetric skip connections," in *Advances in neural information processing systems*, 2016, pp. 2802–2810.

[71] J. Kim, J. Kwon Lee, and K. Mu Lee, "Accurate image super-resolution using very deep convolutional networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 1646–1654.

[72] Z. Cui, H. Chang, S. Shan, B. Zhong, and X. Chen, "Deep network cascade for image super-resolution," in *European Conference on Computer Vision*. Springer, 2014, pp. 49–64.

[73] C. Dong, C. C. Loy, K. He, and X. Tang, "Image super-resolution using deep convolutional networks," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 38, no. 2, pp. 295–307, 2016.

[74] M. Vulović, E. Franken, R. B. Ravelli, L. J. van Vliet, and B. Rieger, "Precise and unbiased estimation of astigmatism and defocus in transmission electron microscopy," *Ultramicroscopy*, vol. 116, pp. 115–134, 2012.

[75] B. Berkels and B. Wirth, "Joint denoising and distortion correction of atomic scale scanning transmission electron microscopy images," *Inverse Problems*, vol. 33, no. 9, p. 095002, 2017.

[76] M. Weigert, U. Schmidt, T. Boothe, M. Andreas, A. Dibrov, A. Jain, B. Wilhelm, D. Schmidt, C. Broaddus, S. Culley *et al.*, "Content-Aware Image Restoration: Pushing the Limits of Fluorescence Microscopy," *bioRxiv*, p. 236463, 2017.

[77] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *in Proceedings of the International conference on machine learning*, 2015, pp. 448–456.

[78] J. Kiefer and J. Wolfowitz, "Stochastic estimation of the maximum of a regression function," *The Annals of Mathematical Statistics*, pp. 462–466, 1952.

[79] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Transactions on image processing*, vol. 13, no. 4, pp. 600–612, 2004.

[80] Z. Wang and A. C. Bovik, "Mean squared error: Love it or leave it? a new look at signal fidelity measures," *IEEE Signal processing magazine*, vol. 26, no. 1, pp. 98–117, 2009.

[81] B. Bajić, J. Lindblad, and N. Sladoje, "Blind restoration of images degraded with mixed poisson-gaussian noise with application in transmission electron microscopy," in *Biomedical Imaging (ISBI), 2016 IEEE 13th International Symposium on*. IEEE, 2016, pp. 123–127.

[82] R. L. Siegel, K. D. Miller, and A. Jemal, "Cancer statistics, 2018," *CA: a cancer journal for clinicians*, vol. 68, no. 1, pp. 7–30, 2018.

[83] M. Malvezzi, G. Carioli, P. Bertuccio, P. Boffetta, F. Levi, C. La Vecchia, and E. Negri, "European cancer mortality predictions for the year 2017, with focus on lung cancer," *Annals of Oncology*, vol. 28, no. 5, pp. 1117–1123, 2017.

[84] "Ers lung cancer facts," https://www.erswhitebook.org/files/public/Chapters/19_lung_cancer.pdf, accessed: 2018-03-24.

[85] "Cancer facts and statistics," https://www.cancer.org/research/cancer-facts-statistics/all-cancer-facts-figures/cancer-facts-figures-2018.html, accessed: 2018-03-24.

[86] G. Brett, "The value of lung cancer detection by six-monthly chest radiographs," *Thorax*, vol. 23, no. 4, pp. 414–420, 1968.

[87] R. S. Fontana, D. R. Sanderson, W. F. Taylor, L. B. Woolner, W. E. Miller, J. R. Muhm, and M. A. Uhlenhopp, "Early lung cancer detection: results of the initial (prevalence) radiologic and cytologic screening in the mayo clinic study," *American Review of Respiratory Disease*, vol. 130, no. 4, pp. 561–565, 1984.

[88] A. Kubik and J. Polak, "Lung cancer detection results of a randomized prospective study in czechoslovakia," *Cancer*, vol. 57, no. 12, pp. 2427–2437, 1986.

[89] J. K. Frost, W. C. Ball Jr, M. L. Levin, M. S. Tockman, R. R. Baker, D. Carter, J. C. Eggleston, Y. S. Erozan, P. K. Gupta, N. F. Khouri *et al.*, "Early lung cancer detection: results of the initial (prevalence) radiologic and cytologic screening in the johns hopkins study," *American Review of Respiratory Disease*, vol. 130, no. 4, pp. 549–554, 1984.

[90] M. R. Melamed, "Lung cancer screening results in the national cancer institute new york study," *Cancer*, vol. 89, no. S11, pp. 2356–2362, 2000.

[91] C. I. Henschke, D. I. McCauley, D. F. Yankelevitz, D. P. Naidich, G. McGuinness, O. S. Miettinen, D. M. Libby, M. W. Pasmantier, J. Koizumi, N. K. Altorki *et al.*, "Early lung cancer action project: overall design and findings from baseline screening," *The Lancet*, vol. 354, no. 9173, pp. 99–105, 1999.

[92] N. L. S. T. R. Team, "The national lung screening trial: overview and study design," *Radiology*, vol. 258, no. 1, pp. 243–253, 2011.

[93] N. L. S. T. R. Team, "Reduced lung-cancer mortality with low-dose computed tomographic screening," *New England Journal of Medicine*, vol. 365, no. 5, pp. 395–409, 2011.

[94] H.-U. Kauczor, L. Bonomo, M. Gaga, K. Nackaerts, N. Peled, M. Prokop, M. Remy-Jardin, O. von Stackelberg, J.-P. Sculier, E. S. of Radiology (ESR *et al.*, "Esr/ers white paper on lung cancer screening," *European radiology*, vol. 25, no. 9, pp. 2519–2531, 2015.

[95] R. J. van Klaveren, M. Oudkerk, M. Prokop, E. T. Scholten, K. Nackaerts, R. Vernhout, C. A. van Iersel, K. A. van den Bergh, S. van't Westeinde, C. van der Aalst *et al.*, "Management of lung nodules detected by volume ct scanning," *New England Journal of Medicine*, vol. 361, no. 23, pp. 2221–2229, 2009.

[96] C. Jacobs, "Automatic detection and characterization of pulmonary nodules in thoracic ct scans," Ph.D. dissertation, [Sl: sn], 2015.

[97] "Computed tomographic image," https://pocketdentistry.com/14-other-imaging-modalities/, accessed: 2018-03-16.

[98] L. E. Romans, *Computed tomography for technologists: Exam review*. Lippincott Williams & Wilkins, 2010.

[99] D. M. Hansell, A. A. Bankier, H. MacMahon, T. C. McLoud, N. L. Muller, and J. Remy, "Fleischner society: glossary of terms for thoracic imaging," *Radiology*, vol. 246, no. 3, pp. 697–722, 2008.

[100] C. I. Henschke, D. F. Yankelevitz, R. Mirtcheva, G. McGuinness, D. McCauley, and O. S. Miettinen, "Ct screening for lung cancer: frequency and significance of part-solid and nonsolid nodules," *American Journal of Roentgenology*, vol. 178, no. 5, pp. 1053–1057, 2002.

[101] K.-I. Kim, C. W. Kim, M. K. Lee, K. S. Lee, C.-K. Park, S. J. Choi, and J. G. Kim, "Imaging of occupational lung disease," *Radiographics*, vol. 21, no. 6, pp. 1371–1391, 2001.

[102] A. McWilliams, M. C. Tammemagi, J. R. Mayo, H. Roberts, G. Liu, K. Soghrati, K. Yasufuku, S. Martel, F. Laberge, M. Gingras *et al.*, "Probability of cancer in pulmonary nodules detected on first screening ct," *New England Journal of Medicine*, vol. 369, no. 10, pp. 910–919, 2013.

[103] A. A. Setio, C. Jacobs, J. Gelderblom, and B. Ginneken, "Automatic detection of large pulmonary solid nodules in thoracic ct images," *Medical Physics*, vol. 42, no. 10, pp. 5642–5653, 2015.

[104] S. G. Armato, R. Y. Roberts, M. Kocherginsky, D. R. Aberle, E. A. Kazerooni, H. MacMahon, E. J. van Beek, D. Yankelevitz, G. McLennan, M. F. McNitt-Gray *et al.*, "Assessment of radiologist performance in the detection of lung nodules: dependence on the definition of "truth"," *Academic radiology*, vol. 16, no. 1, pp. 28–38, 2009.

[105] J. K. Leader, T. E. Warfel, C. R. Fuhrman, S. K. Golla, J. L. Weissfeld, R. S. Avila, W. D. Turner, and B. Zheng, "Pulmonary nodule detection with low-dose ct of the lung: agreement among radiologists," *American Journal of Roentgenology*, vol. 185, no. 4, pp. 973–978, 2005.

[106] G. D. Rubin, J. K. Lyo, D. S. Paik, A. J. Sherbondy, L. C. Chow, A. N. Leung, R. Mindelzun, P. K. Schraedley-Desmond, S. E. Zinck, D. P. Naidich *et al.*, "Pulmonary nodules on multi–detector row ct scans: performance comparison of radiologists and computer-aided detection," *Radiology*, vol. 234, no. 1, pp. 274–283, 2005.

[107] F. Beyer, L. Zierott, E. Fallenberg, K. Juergens, J. Stoeckel, W. Heindel, and D. Wormanns, "Comparison of sensitivity and reading time for the use of computer-aided detection (cad) of pulmonary nodules at mdct as concurrent or second reader," *European radiology*, vol. 17, no. 11, pp. 2941–2947, 2007.

[108] M. L. Giger, H.-P. Chan, and J. Boone, "Anniversary paper: History and status of cad and quantitative image analysis: the role of medical physics and aapm," *Medical physics*, vol. 35, no. 12, pp. 5799–5820, 2008.

[109] A. Christe, L. Leidolt, A. Huber, P. Steiger, Z. Szucs-Farkas, J. Roos, J. Heverhagen, and L. Ebner, "Lung cancer screening with ct: evaluation of radiologists and different computer assisted detection software (cad) as first and second readers for lung nodule detection at different dose levels," *European journal of radiology*, vol. 82, no. 12, pp. e873–e878, 2013.

[110] G. Iussich, L. Correale, C. Senore, N. Segnan, A. Laghi, F. Iafrate, D. Campanella, E. Neri, F. Cerri, C. Hassan *et al.*, "Ct colonography: preliminary assessment of a double-read paradigm that uses computer-aided detection as the first reader," *Radiology*, vol. 268, no. 3, pp. 743–751, 2013.

[111] R. Hupse, M. Samulski, M. Lobbes, A. Den Heeten, M. W. Imhof-Tas, D. Beijerinck, R. Pijnappel, C. Boetes, and N. Karssemeijer, "Standalone computer-aided detection compared to radiologists' performance for the detection of mammographic masses," *European radiology*, vol. 23, no. 1, pp. 93–100, 2013.

[112] K. Doi, "Computer-aided diagnosis in medical imaging: historical review, current status and future potential," *Computerized medical imaging and graphics*, vol. 31, no. 4-5, pp. 198–211, 2007.

[113] M. Silva, C. Schaefer-Prokop, C. Jacobs, G. Capretti, F. Ciompi, B. van Ginneken, U. Pastorino, and N. Sverzellati, "Detection of subsolid nodules in lung cancer screening: Complementary sensitivity of visual reading and computer-aided diagnosis." *Investigative radiology*, 2018.

[114] M. Prokop, "Lung cancer screening: the radiologist's perspective," in *Seminars in respiratory and critical care medicine*, vol. 35, no. 01. Thieme Medical Publishers, 2014, pp. 091–098.

[115] E. Lopez Torres, E. Fiorina, F. Pennazio, C. Peroni, M. Saletta, N. Camarlinghi, M. Fantacci, and P. Cerello, "Large scale validation of the m5l lung cad on heterogeneous ct datasets," *Medical physics*, vol. 42, no. 4, pp. 1477–1489, 2015.

[116] M. S. Brown, P. Lo, J. G. Goldin, E. Barnoy, G. H. J. Kim, M. F. McNitt-Gray, and D. R. Aberle, "Toward clinically usable cad for lung cancer screening with computed tomography," *European radiology*, vol. 24, no. 11, pp. 2719–2728, 2014.

[117] M. Tan, R. Deklerck, J. Cornelis, and B. Jansen, "Phased searching with neat in a time-scaled framework: experiments on a computer-aided detection system for lung nodules," *Artificial intelligence in medicine*, vol. 59, no. 3, pp. 157–167, 2013.

[118] L. Lu, Y. Tan, L. H. Schwartz, and B. Zhao, "Hybrid detection of lung nodules on ct scan images," *Medical physics*, vol. 42, no. 9, pp. 5042–5054, 2015.

[119] T. Messay, R. C. Hardie, and S. K. Rogers, "A new computationally efficient cad system for pulmonary nodule detection in ct imagery," *Medical image analysis*, vol. 14, no. 3, pp. 390–406, 2010.

[120] S. G. Armato, M. L. Giger, and H. MacMahon, "Automated detection of lung nodules in ct scans: preliminary results," *Medical physics*, vol. 28, no. 8, pp. 1552–1561, 2001.

[121] Q. Dou, H. Chen, L. Yu, J. Qin, and P.-A. Heng, "Multilevel contextual 3-d cnns for false positive reduction in pulmonary nodule detection," *IEEE Transactions on Biomedical Engineering*, vol. 64, no. 7, pp. 1558–1567, 2017.

[122] J.-Z. Cheng, D. Ni, Y.-H. Chou, J. Qin, C.-M. Tiu, Y.-C. Chang, C.-S. Huang, D. Shen, and C.-M. Chen, "Computer-aided diagnosis with deep learning architecture: applications to breast lesions in us images and pulmonary nodules in ct scans," *Scientific reports*, vol. 6, p. 24454, 2016.

[123] A. A. Setio, F. Ciompi, G. Litjens, P. Gerke, C. Jacobs, S. J. van Riel, M. M. W. Wille, M. Naqibullah, C. I. Sánchez, and B. van Ginneken, "Pulmonary nodule detection in ct images: false positive reduction using multi-view convolutional networks," *IEEE Transactions on medical imaging*, vol. 35, no. 5, pp. 1160–1169, 2016.

[124] S. G. Armato, G. McLennan, L. Bidaut, M. F. McNitt-Gray, C. R. Meyer, A. P. Reeves, B. Zhao, D. R. Aberle, C. I. Henschke, E. A. Hoffman *et al.*, "The lung image database consortium (lidc) and image database resource initiative (idri): a completed reference database of lung nodules on ct scans," *Medical physics*, vol. 38, no. 2, pp. 915–931, 2011.

[125] C. I. Henschke, D. P. Naidich, D. F. Yankelevitz, G. McGuinness, D. I. McCauley, J. P. Smith, D. Libby, M. Pasmantier, M. Vazquez, J. Koizumi *et al.*, "Early lung cancer action project: initial findings on repeat screenings." *Cancer*, vol. 92, no. 1, pp. 153–159, 2001.

[126] S. G. Armato, L. M. Hadjiiski, G. D. Tourassi, K. Drukker, M. L. Giger, F. Li, G. Redmond, K. Farahani, J. S. Kirby, and L. P. Clarke, "Special section guest editorial: Lungx challenge for computerized lung nodule classification: reflections and lessons learned," *Journal of Medical Imaging*, vol. 2, no. 2, p. 020103, 2015.

[127] M. Niemeijer, M. Loog, M. D. Abramoff, M. A. Viergever, M. Prokop, and B. van Ginneken, "On combining computer-aided detection systems," *IEEE Transactions on Medical Imaging*, vol. 30, no. 2, pp. 215–223, 2011.

[128] R. Manniesing, M. A. Viergever, and W. J. Niessen, “Vessel enhancing diffusion: A scale space representation of vessel structures,” *Medical image analysis*, vol. 10, no. 6, pp. 815–825, 2006.

[129] K. Lidayová, “Fast methods for vascular segmentation based on approximate skeleton detection,” Ph.D. dissertation, Acta Universitatis Upsaliensis, 2017.

[130] K. Lidayova, H. Frimmel, C. Wang, E. Bengtsson, and Ö. Smedby, “Fast vascular skeleton extraction algorithm,” *Pattern Recognition Letters*, vol. 76, pp. 67–75, 2016.

ACKNOWLEDGEMENTS

The past four years have been one of the most challenging yet exciting journeys for me, and I feel very grateful to experience it first hand. I couldn't have completed this doctoral research without the support of many people. I would like to take this opportunity to recognize and commemorate in print, my thanks to those who have helped me throughout my studies.

I had the pleasure of working in two wonderful environments thanks to my colleagues and head of the T. J. Seebeck Dept. of Elec., *Prof. Thomas Rang*, as well as my friends, colleagues, and head of Centre of Image Analysis, *Prof. Ingela Nyström*.

I thank *Prof. Olev Märtens*, my main supervisor, for the opportunity to conduct my Ph.D. research and the advice, encouragement, and support over my Ph.D. years.

My sincerest appreciation also goes to *Prof. Yannick Le Moullec*, my co-supervisor, for his guidance, dedication, and motivation towards my research work. Thank you for being a great advisor.

I would like to express my deepest gratitude to my dearest supervisor, *Assoc. Prof. Ida-Maria Sintorn*, for her encouragement, criticism, and love during the most critical period of my Ph.D. journey. I am especially grateful for the pivotal role she played when she joined my supervising team or even before it. Your great enthusiasm for the research and philosophy of life have enlightened me in a significant way. I always look up to you not just being a mentor, but a role model, and a great friend forever.

I would like to express great appreciation to *Dr. Tõnis Saar*, my external consultant, for the invaluable support and guidance throughout my research. Thank you for being very supportive.

My collaborators, *Buda Bajić*, *Kristina Lidayová*, *Nataša Sladoje*, *Anca Dragomir*, *Ivana Pepić*, *Amit Suveer*, *Joakim Lindblad*, *Ewert Bengtsson*, *Örjan Smedby*, and *Hans Frimmel*, for fruitful discussions, good collaboration and valuable opinions, and comments.

Prof. Mart Min, *Prof. Enn Velmre*, *Ida-Maria Sintorn*, *Ingela Nyström*, *Carolina Wählby*, *Nataša Sladoje*, *Buda Bajić*, *Kristina Lidayová*, *Eva Breznik*, *Leslie Solorzano*, *Joakim Lindblad*, *Amit Suveer*, and *Johan Öfverstedt*, for proofreading and commenting on my thesis. *Damian Matuszewski* for your valuable suggestions in thesis structuring.

My colleagues/friends in Uppsala, *Kristina Lidayová*, *Eva Breznik*, *Giorgia*, *Ekta Vats*, *Leslie Solorzano*, *Elisabeth Wetzer*, *Amit Suveer*, *Rameez Malik*, *Damian Matuszewski*, *Sajith Sadanandan*, *Maxime Bombrun*, *Christophe Avenel*, *Nadia Assraoui*, *Kalyan Ram*, *Teo Asplund*, *Johan Öfverstedt*, and *Fredrik Nysjö*. My colleagues/friends in Tallinn, *Hip Kõiv*, *Jane Rang*, *Eva Keerov*, *Tauseef Ahmed*, *Faisal Ahmed*, *Fredrik Rang*, and *Alvar Kurrel*. aitäh *Alvar* for teaching me a bit of Estonian. :-)

My good friends in Tallinn, *Jawad*, *Arbind*, *Aditya*, *Junaid*, *Qasim*, *Jaan*, and *Geiu* for being caring and respectful. Good luck for your bright future.

Joakim Lindblad for giving me an insight into what, why, and how. I sincerely appreciate your critical thinking.

Nataša Sladoje for your appreciation, cheerfulness and good humor. I hope that I manage to convince you about deep learning since our first discussion at ISBI'16, Prague. :-)

Damian Matuszewski and *Sabine Radde*, for your love, support, and care during my visits to Uppsala. Best wishes for the new family member! :-)

Amit Suveer, for being a good friend, and very supportive throughout last three years. Thanks for the nice and pleasant time during my visits to Uppsala.

My dearest friend, *Buda Bajić*, for your patience and sincerity while teaching me a bit of mathematics. I really enjoyed the joint nights while working alongside you on our ISBI'18 paper. Also, thanks for being a great host and showing me around Novi Sad. :-)

Mērī píárī *Jānā*, tusīṁ kīmatī hō atē mērē láī sabha tōṁ vadhīā cīzāṁ vicōṁ ika hō. Maiṁ tuhānū bahuta píára karadā hāṁ atē hamēśā láī píára karēgā. Tuhāḍē samarathana, nēṛē hōṇa, atē āpaṇī sārī zidagī jī́uṇa láī mainū śānadāra yādāṁ dēṇa láī tuhāḍā dhanavāda.

My sister *Anjana*, for being the best sister one could wish for. Best wishes to you for your married life. My father *Arvind*, for letting me fly, constant support and encouragements that helped me throughout my studies. My dearest Uncle, *Umesh*, for being so kind to me. Thank a lot for your love and support. I have learned a lot about life from you. My aunt, *Suman*, for being so caring. *Puru*, my cute little brother, welcome to our family. Loads of love and looking forward to meeting you in person.

Although I have dedicated this thesis to my grandfather, it is equally devoted to my mother *Kamlesh* as well. Thank you for the sacrifices you have made to give me the life I have. I miss you a lot and always wish for your presence.

I am thankful to the following entities for their financial support during my PhD studies:

- Estonian National Scholarship Programme for International Students, Researchers and Academic staff.
- IT Academy Scholarship for PhD Students of Information and communication Technology.

ABSTRACT
Classification and Denoising of Objects in TEM and CT Images Using Deep Neural Networks

The digitization of biomedical and medical images has benefited the clinicians in comprehending (or detecting) obscure abnormalities. However, manual analysis is labor-intensive and time-consuming. Since the last few decades, computer-aided detection (CAD) systems employing learning-based methods and conventional image analysis-based methods have successfully paved the landscape for the detection (and/or classification) of deadly abnormalities. Lately, the inception of deep neural networks (DNN) (often synonymized as deep learning) as a powerful recognition module has shifted the research interest from problem-specific solutions to increasingly problem-agnostic methods that rely on learning from data. In particular, convolutional neural networks (CNNs) have rapidly become a primary choice for many CAD systems due to their astonishing results. This impulse has been sparked by increased computational power (graphical processing units) and the evolution of learning-based methods.

This thesis presents a total of five solutions: four DNN-based solutions for classification of structures in biomedical and medical images, and one solution for denoising of biomedical images to improve the image quality. This thesis is focused on the applications of two variants of DNN: the CNN, and the multi-layer perceptrons (MLP).

From a biomedical image analysis perspective, the first solution is associated with improving the performance of automated workflow for primary ciliary dyskinesia (PCD) analysis. To *classify* cilia and non-cilia structures in low-magnification (LM) transmission electron microscopy (TEM) images, a CNN-based classifier is developed as a false positives (FP) reduction module. Although computing discriminative features of cilia structures at very low magnification is challenging, the developed CNN classifier substantially improves the F-score from 0.47 to 0.59.

The second solution takes a side step from classification and focuses on denoising. Denoising is often considered as a preprocessing step to improve the image quality for automated analysis. Given this, the second solution is associated with enhancing the structural information in short exposure high-magnification TEM images. A novel multi-stream CNN-based model is developed to *denoise* 100 short exposure HM images acquired at the same spatial location in the cell

section. Three different strategies for combining denoising and image merging are investigated to determine the optimal structure enhancing strategy. The CNN denoising model is only trained for one strategy and used as it is for other two strategies, thus presenting the transfer learning perspective of DNN as a potential add-on to automated analysis. The presented model achieves an improved PSNR of 40.84 dB.

From a medical image analysis perspective, the third solution is associated with improving the performance of a CAD system for the early detection of multiple sizes of nodules (3 - 30 mm) in computed tomography (CT) scans. To *classify* nodules and non-nodules, an MLP-based classifier is developed as an FP reduction module. The CAD is extensively tested on four publically available CT datasets; this makes it the only system to be successfully validated on such large scale. The developed CAD system achieves a high sensitivity of 85.6% with only 8 FPs/scan.

Until recently, conventional CAD systems employing learning-based methods depended on handcrafted representations (features). Designing features by hand is challenging and often result in limited discriminative power; thus, this is insufficient to classify micronodules (≤ 4 mm) and cross-sectional vessels. The fourth solution is associated with developing a CAD system for the detection of micronodules in CT scans. To *classify* micronodules and small cross-sectional vessels, a novel 3D CNN classifier is developed as an FP reduction module. Using the largest publically available CT dataset, the developed CAD system achieves a high sensitivity of 86.7% with only 8 FPs/scan.

The fifth solution is associated with improving the performance and efficiency of automated workflow for detecting multiple sizes of vascular nodes in CT angiography (CTA) scans. To *classify* cross-sections of different sizes of vessel and non-vessel nodes, a patch-based CNN classifier is developed as an FP reduction module. On the given 25 CTA volumes from the clinical routine, the presented classifier substantially improves the F-score from 0.43 to 0.82.

KOKKUVÕTE
Objektide klassifitseerimine ja müratustamine TEM ja KT kujutistelt sügavate närvivõrkude abil

Meditsiiniliste ja biomeditsiinitehniliste kujutiste digiteerimine on kliiniliselt abiks võimalike ebanormaalsuste mõistmisel (või avastamisel). Potensiaalsed võimalused realiseeruvad aga suure hulgal pildiandmete töötlusel. Samas on kujutiseandmete hulga käsitsi tõlgendamine töömahukas ja aeganõudev. Arvestades andmete rohkust kujutiste eri modaalsustest, püüavad teadlased arendada automaatseid analüüsimeetodeid või arvutipõhist avastamist (CAD), et abistada arste selles monotoonses töös. Viimastel aastakümnetel on automatiseeritud biomeditsiiniliste (ja meditsiiniliste) piltide analüüs edukalt sillutanud teed CAD-süsteemidele, mis kasutavad masinõppepõhiseid meetodeid ja tavapärast pildianalüüsi. Kõigi õpipõhiste meetodite seas on närvivõrgud (NN-d) suutlikud tuvastama (ja/või klassifitseerida) eluohtlikke kõrvalekaldeid ja omavad klassifitseerimiseks suurt potentsiaali. Kuid CAD-süsteemide toimivus on viimase kümne aasta jooksul olnud piiratud mitteküllaldase arvutusvõimsuse tõttu. Viimasel ajal on sügavate närvivõrkude (sügavõppe sünonüümina) loomine võimsaks tuvastusmooduliks. Probleemipõhistest lahendustest pärineva uurimistöö huvi on järjest enam nihkunud probleemi-agnostilistele lahendustele, mis põhinevad andmetest õppimisel. Eriti on konvolutsioonilised närvivõrgud (CNN-d) muutunud esmaseks valikuks paljude CAD-süsteemide puhul nende suurepäraste tulemuste tõttu. Seda arengut on põhjustanud arvutusvõimsuse suurenemine, eriti graafikaprotsessorite (GPU) ja õpimeetodite arengu tõttu.

Arvestades DNNide prevaleerumist, esitatakse käesolevas doktoritöös kokku viis lahendust: neli DNN-põhist lahendust struktuuride klassifitseerimiseks nii biomeditsiinilistelt kui ka meditsiinilistelt piltidelt ja üks lahendus on biomeditsiiniliste kujutiste müratustamiseks, et parendada pildikvaliteeti. See väitekiri keskendub DNN-ide kahe variandi - mitmekihiliste pertseptronite (MLP) ja CNN-ide - rakendustele.

Biomeditsiinilise pildianalüüsi perspektiivist on esimene lahendus seotud primaarse tsiliaarse düskineesia (PCD) analüüsi automatiseeritud töövoo täiustamisega. Väikse suurendusega (LM) transmissioon-elektronmikroskoopia (TEM) piltidelt tsiiliate ja mitte-tsiiliate klassifitseerimisel on valepositiivse (FP) vähendusmoodulina välja töötatud CNN-põhine klassifikaator. Kuigi tsiiliate omadusi saab vaevu eristada, arvutamaks diskrimineerivaid funktsioone väga

väikese kujutise suurendusega, parendab CNN klassifikaator oluliselt F-skoori (0,47-0,59-ni).

Teine lahendus liigub klassifitseerimiselt kõrvale ja keskendub kujutiste müratustamisele. Pildikvaliteedi parendamist müratustamisega peetakse automaatsel analüüsil tihti eelprotsessiks. Seda arvestades on teine pakutud lahendus seotud struktuurialase teabe täiustamisega MiniTEM™-i abil omandatud lühikese säritusega suure suurendusega (HM) TEM-kujutistel. Uuendusliku mitmekanalilise CNN-põhise mudeli väljatöötamisel võetakse 100 vähesäritatud HM-pilti, mis on saadud raku sektsiooni samas ruumilises asukohas. Optimaalse müratustamisstrateegia kindlaksmääramiseks uuritakse kolme erinevat strateegiat. Mudelit on õpetatud ainult ühe strateegia jaoks ja seda kasutatakse ka kahe teise strateegia puhul; esitades seega DNN-ide ülekandeõppe perspektiivi automatiseeritud analüüsi võimaliku lisana. Esitatav mudel saavutab parema PSNRi (signaal-mürasuhte tippväärtus) - kuni 40,84 dB.

Meditsiinilise kujutise analüüsi vaatepunktist on kolmas lahendus seotud kompuutertomograafia (CT) skaneerimisega eri suurusega noodulite (3 ... 30 mm) varajase avastamise CAD süsteemi täiustamisega. Noodulite ja mittenoodulite efektiivseks liigitamiseks on vale-positiivsete (FP) vähendamise moodulina välja töötatud MLP-põhine klassifikaator. CAD on põhjalikult testitud nelja avalikult kättesaadava CT-andmestikuga; seega tegemist on ainukese CAD-süsteemiga, mida on sellisel suurel määral edukalt valideeritud. Arendatud CAD-süsteem saavutab parema tundlikkuse (kuni 85,6%) vaid 8 valepositiivse tulemusega skänni kohta.

Kuni viimase ajani sõltusid tavapärased CAD-süsteemid, mis kasutasid õpipõhiseid meetodeid, käsitsi ettevalmistatud esitustest (funktsioonidest). Funktsioonide käsitsi kavandamine on tülikas ja sageli on piiratud diskrimineeriv jõudlus; seega pole see mikronoodulite (<4 mm) ja ristlõikeliste veresoonte liigitamiseks piisav. Neljas lahendus on seotud CAD-süsteemi väljatöötamisega mikronoodulite tuvastamiseks CT-skaneerimisel. Mikronoodulite ja väikese ristlõikega veresoonte klassifitseerimiseks on FP vähendusmoodulina välja töötatud uudne 3D-CNN klassifikaator. Kasutades suurimat avalikult kättesaadavat CT-andmestikku võib väita, et arendatud CAD-süsteem on kõrge tundlikkusega (86,7%), vaid 8 valepositiivsega skänni kohta.

Meditsiinilise kujutise analüüsi vaatepunktist on lõplik lahendus seotud automatiseeritud töövoo efektiivsuse ja tõhususe parendamisega CT angiograafia (CTA) skaneerimisel arvukate veresoonte sõlmede tuvastamiseks. Veresoonte ja mitte-veresoonte sõlmede erineva suurusega ristlõigete klassifitseerimiseks on FP vähendusmooduliks paketipõhine CNN klassifikaator. Antud kahekümne viie CTA kliinilisest rutiinist annab CNN-põhiste klassifikaator oluliselt parendatud F-skoori (0,43 asemel 0,82).

APPENDIX

Publication A

Appeared in:

Gupta A, Suveer A, Lindblad J, Dragomir A, Sintorn I-M, and Sladoje N.

Convolutional Neural Networks for False Positive Reduction of Automatically Detected Cilia in Low Magnification TEM Images. In *Proceeding of the 20th Scandinavian Conference on Image Analysis (SCIA)*, Tromsø, Norway, June 2017, LNCS-10269, pp 407-418. doi: /10.1007/978-3-319-59126-1_34

Convolutional Neural Networks for False Positive Reduction of Automatically Detected Cilia in Low Magnification TEM Images

Anindya Gupta[1(✉)], Amit Suveer[2], Joakim Lindblad[2,3], Anca Dragomir[4], Ida-Maria Sintorn[2,5], and Nataša Sladoje[2,3]

[1] T.J. Seebeck Department of Electronics,
Tallinn University of Technology, Tallin, Estonia
anindya.gupta@ttu.ee
[2] Department of IT, Centre for Image Analysis,
Uppsala University, Uppsala, Sweden
{amit.suveer,joakim.lindblad,
ida.sintorn,natasa.sladoje}@it.uu.se
[3] Mathematical Institute,
Serbian Academy of Sciences and Arts, Belgrade, Serbia
[4] Department of Surgical Pathology,
Uppsala University Hospital, Uppsala, Sweden
anca.dragomir@igp.uu.se
[5] Vironova AB, Stockholm, Sweden

Abstract. Automated detection of cilia in low magnification transmission electron microscopy images is a central task in the quest to relieve the pathologists in the manual, time consuming and subjective diagnostic procedure. However, automation of the process, specifically in low magnification, is challenging due to the similar characteristics of non-cilia candidates. In this paper, a convolutional neural network classifier is proposed to further reduce the false positives detected by a previously presented template matching method. Adding the proposed convolutional neural network increases the area under Precision-Recall curve from 0.42 to 0.71, and significantly reduces the number of false positive objects.

Keywords: Convolutional neural network · Primary Ciliary Dyskinesia · Template maching · Transmission electron microscopy

1 Introduction

Primary Ciliary Dyskinesia (PCD) is a rare genetic disorder resulting in dysfunctional cilia - the hairlike structures protruding from certain cells. Dysfunctionality of cilia can result in severe chronic respiratory infection, and infertility in both genders. To diagnose the disorder, pathologists examine the morphological appearance of cilia (∼220–250 nm) using transmission electron microscopy

P. Sharma and F.M. Bianchi (Eds.): SCIA 2017, Part I, LNCS 10269, pp. 407–418, 2017.
DOI: 10.1007/978-3-319-59126-1_34

(TEM). Qualitative analysis of cilia in the TEM images is still largely subjective and manual diagnosis is laborious, monotonous, and hugely time consuming (diagnosis takes ca. two hours per sample). An expert pathologist has to zoom in and out at locations of cilia which possibly exhibit structural information necessary for correct diagnosis. Navigation through the huge search space, together with change of magnification, is very demanding. Hence, there is an inevitable requisite for the automation of the cilia detection and diagnosis process. However, it is not feasible to acquire images which cover the whole sample at a magnification that allows structural analysis; such an acquisition would take tens of hours. Furthermore, objects of interest are rare, very small, and not spreading over more than a couple of percents of the total sample. Locating these regions of interest at low magnification, and acquiring high magnification images only at selected locations, would therefore be highly beneficial.

Automated detection of cilia structures (of a quality sufficient for diagnosis) at low magnification is a challenging task due to (1) their similar characteristics with the large number of non-cilia structures, and (2) variance in the size, shape and appearance of the individual cilia structures. The task becomes more complicated also due to noise and the non-homogeneous background at low magnification, see Fig. 1.

Lately, availability of large amounts of data and strong computational power have rapidly increased the popularity of machine learning approaches (deep learning). Convolutional neural networks (CNN) [10] have outperformed the state-of-the-art in many computer vision applications [8]. Similarly, the applicability of

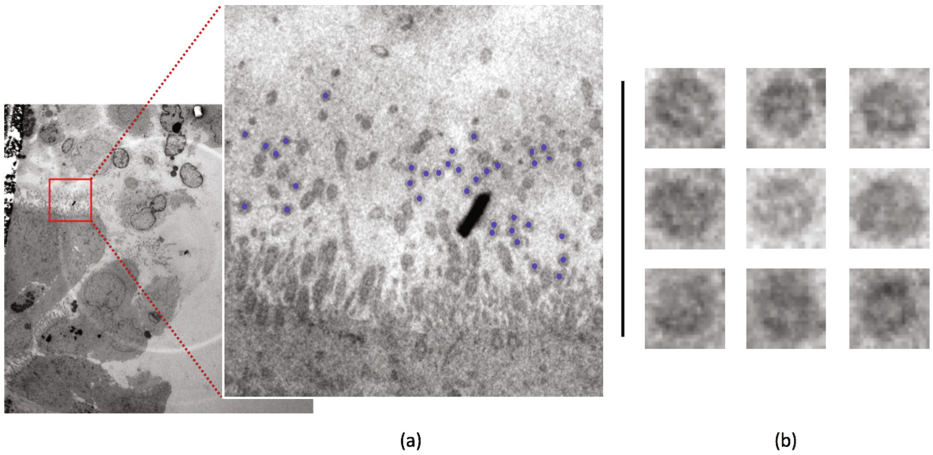


Fig. 1. (a) Low magnification TEM image of 4096×4096 pixels utilized for training purpose with the magnified view of 350×350 pixel bounding box (marked in red) with indicated ground truth marked by an expert pathologist. Here, cilia candidates marked with blue dots are of the suitable quality. (b) Some examples of patches extracted by previously reported method [15], the first and second rows contain true positives (TP) whereas patches in the third row are false positives (FP). Note the high similarity between the classes, this makes the problem a serious challenge. (Color figure online)

CNN is also investigated in the medical image analysis field [1,11]. In particular, their capability to learn discriminative features while trained in a supervised fashion makes them useful for automated detection of structures in, e.g., electron microscopy images. For instance, Ciresan *et al.* [5] reported a CNN model to segment the neuronal membranes in electron microscopy images; in [19], a CNN with autoencoder for automated detection of nuclei in high magnification (HM) microscopy images was employed.

Previously, a template matching (TM) method to detect cilia candidates in low magnification TEM images was proposed [15]. Considering that we aim at locating regions highly populated by good quality cilia, for further HM image acquisition and analysis, it is crucial that the identification of such regions is not misled by a large number of false positives (FP). In the current work, we aim at improving the performance by incorporating a dedicated CNN model in the cilia detection scheme with the special focus on reducing the number of FP. A performance benchmark for the proposed model is presented, and independent validation on an additional image is performed.

2 Image Data

Two low magnification (LM) TEM images from different patients, each with ca. 200 cilia structures, are used for training and independent validation purposes. Both images are acquired with a FEI Tecnai G2 F20 TEM and a bottom mounted FEI Eagle 4K × 4K HR CCD camera, resulting in 16-bit gray scale TIFF images of size 4096×4096 pixels.

For each LM image field, a set of mid magnification (MM) images are acquired, where the ground truth, i.e., true cilia candidates of promising quality for diagnosing at HM (not dealt with in this paper), are manually marked by an expert pathologist (author AD). Some examples of extracted patches of marked cilia candidates are shown in Fig. 1(b). The field of view (FOV) for a MM (2900×) image is 15.2 µm and for a LM (690×) image, it is 60.6 µm.

3 Method

The overall detection workflow consists of two stages: (1) Template matching as described in [15], and (2) further FP reduction using a 2-D CNN model, which is the core of this paper.

3.1 Initial Candidate Detection

Template matching based on normalized cross-correlation (NCC) and a customized synthetic template is used to detect the initial cilia candidates. The cross correlation image is thresholded at a suitable threshold, followed by area filtering and position filtering, meaning that only the best hit in a local region is kept as a candidate [15].

3.2 Data Partitioning and Augmentation

For each candidate position, we extracted patches of 23×23 pixels centered at a given position $p = (x, y)$. The patch size was chosen in order to contain a cilia object (∼19–20 pixels diameter), and some local background around the cilia instances (∼3 pixels) to include sufficient context information.

A training set of cilia, as well as non-cilia candidates, was extracted from the training image based on ground truth markings made by our expert pathologist (author AD), in MM images covering the same area of the sample. All true cilia (a total of 136) regardless of their match score, i.e., their NCC values, were chosen. A set of 272 non-cilia candidates was extracted from different NCC levels in order to represent non-cilia objects with high similarity to good cilia (136 randomly chosen non-cilia objects with NCC values ≥ 0.5) as well as non-cilia objects more different from true cilia (136 randomly chosen objects with NCC threshold values between 0.2 and 0.5).

While training a CNN model, an imbalanced dataset can mislead the optimization algorithm to converge to a local minimum, wherein the predictions can be skewed towards the candidates of the majority class, resulting in an overfitted model. To avoid overfitting, candidates from both classes (i.e. cilia and non-cilia) are augmented. Augmentation on test data has shown a considerable improvement in terms of robustness of the system, as it, if designed properly for the problem at hand [3].

Prior to the augmentation step, the candidates are randomly divided into training, validation and test sets. The training set consists of 82 cilia and 164 non-cilia candidates whereas the validation and test sets, each consists of 27 cilia and 54 non-cilia candidates. The candidates are augmented using affine transformations (rotation, scaling and shear) and bilinear interpolation. Horizontal flipping is applied to the cilia candidates to balance the sets. A fully automated script is created to perform the combination of seven random angular rotations (0–360°), six random scalings within ±10% range and five random shearings within 5% range in both x- and y- directions, resulting in 1050 augmented variations for each candidate. The augmentation scheme is applied separately for each subset to ensure independency of the training set from the validation and test sets.

3.3 2-D CNN Configuration

The architecture of the proposed CNN model is initially derived from the LeNet architecture [9]. The motivation behind this choice is its efficiency, as well as lower computational cost compared to the architectures such as Alexnet [8] and VGGnet [13]. These models have extended the functionality of LeNet into a much larger neural network with often better performance but at a cost of a massive increase in number of parameters and computational time. Training of such large networks is still difficult due to the lack of powerful ways to regularize the models and large feature sizes in many layers [16]. Hence, we decided to empirically modify the LeNet architecture to fit our application.

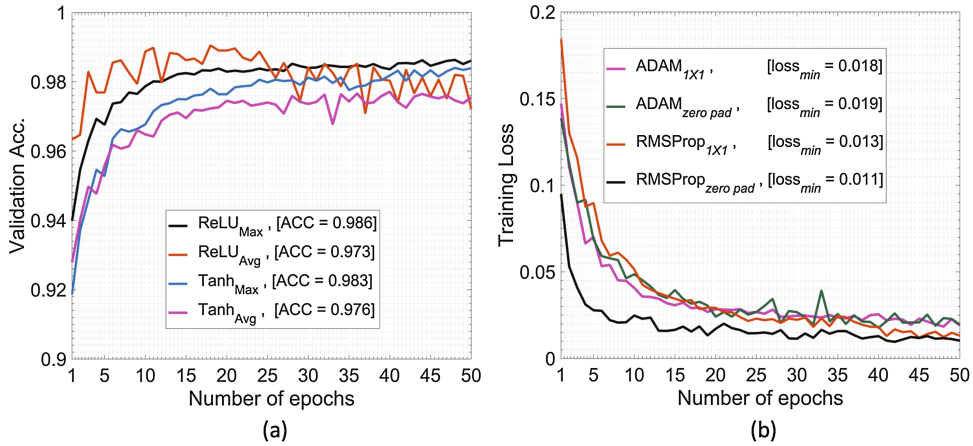


Fig. 2. Performance curves of different configuration: (a) validation accuracy for different activation functions and pooling layer combinations; (b) training loss for different optimizers with zero-padding and kernel of 1×1.

In our modified architecture, the default activation function i.e., hyperbolic tangent (tanh) [18] is replaced with Rectified linear units (ReLU) [12]. In comparison to the tanh, the constant gradient of ReLUs results in faster learning and also reduces the problem of vanishing gradient. We also implemented the maxpooling layer instead of average pooling as subsampling layer [8]. A comparative performance of both activation functions with different subsampling layers are shown in Fig. 2(a). The figure shows the accuracy for each configuration at different number of epochs. It is noticeable that the performance is better when ReLU was configured with maxpooling layer, resulting in higher accuracy after 50 epochs.

We also compared the usability of zero-padding and 1×1 convolution filters (as suggested in [16]) for two different optimizers, Adam [7] and RMSProp [17]. A kernel of size 1×1 in the first convolutional layer reduced the number of parameters (difference of 1 120 parameters compared to the zero-padding), thus keeping the computations reasonable. Comparatively, in either configuration, RMSProp with zero-padding resulted in a better training loss, as shown in Fig. 2(b). We thus, selected the configuration with minimum training loss. Moreover, several parameters (number of layers, kernel size, training algorithm, and number of neurons in the dense layer) were also experimentally determined.

In the proposed CNN classifier, the input patches are initially padded with a three pixels thick frame of zeros in order to keep the spatial sizes of the patches constant after the convolutional layers, as well as to keep the border information up to the last convolutional layer. Next, two consecutive convolutional layers and subsampling layers are used in the network. The first convolutional layer consists of 32 kernels of size $6 \times 6 \times 1$. The second convolutional layer consists of 48 kernels of size $5 \times 5 \times 32$. The subsampling layer is set as the maximum values in non-overlapping windows of size 2×2 (stride of 2). This reduces the size of

the output of each convolutional layer by half. The last layer is a fully connected layer with 20 neurons followed by a softmax layer for binary classification. ReLU are used in the convolutional and dense layers, where the activation y for a given input x is obtained as $y = max\ (0,\ x)$. The architecture of the proposed CNN model is shown in Fig. 3.

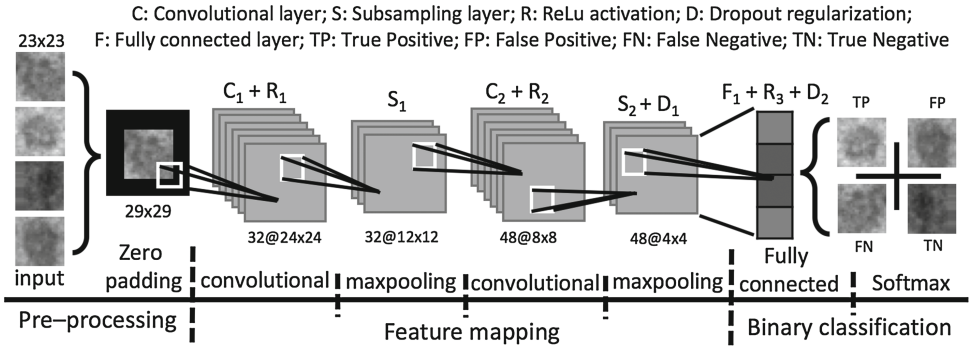


Fig. 3. An overview of the proposed CNN model.

3.4 Network training

The training of the classifier was performed in a 5-fold cross-validation scheme. For each fold, the candidates were randomly split into five blocks to ensure that each set was utilized as test set once. The distribution of candidates in each fold was kept as shown in Table 1.

Table 1. The number of cilia and non-cilia candidates in the different sets. Candidates marked in bold are finally utilized for building the model.

Set	Training	Validation	Test
Cilia	82	27	27
Aug (cilia)	172 364	56 754	56 754
Non-cilia	164	54	54
Aug (non-cilia)	172 364	56 754	56 754
Final set	**344 728**	**113 508**	**113 508**

On the given training dataset, RMSProp [17] is used to efficiently optimize the weights of the CNN. RMSProp is an adaptive optimization algorithm, which normalizes the gradients by utilizing the magnitude of recent gradients. The weights are initialized using normalized initialization as proposed in [6] and updated in a mini-batch scheme of 128 candidates. The biases were initialized with zero and learning rate was set to 0.001. A dropout of 0.5 is implemented

as regularization, on the output of the last convolutional layer and the dense layer to avoid overfitting [14]. Softmax loss (cross-entropy error loss) is utilized to measure the error loss. The CNN model is implemented using theano backend in Keras [4]. The average training time is approximately 48 s/epoch on a GPU GeForce GTX 680.

4 Experimental Results and Discussion

The performance of the proposed CNN model was evaluated in terms of *Precision*, *Recall*, *Area under the Precision-Recall curve (AUC)*, and *F-score*, defined as:

$$Precision = \frac{TP}{TP + FP}, \qquad Recall = \frac{TP}{TP + FN},$$

$$F\text{–}score = 2 \times \frac{Precision \times Recall}{Precision + Recall}, \qquad AUC = \int_0^1 P(r)dr.$$

The AUC is the average of precision $P(r)$ over the interval $(0 \leq r \leq 1)$, and $P(r)$ is a function of recall r. Additionally, for different NCC threshold levels, the Free-response Receiver Operating Characteristic (FROC) curve [2] was utilized to measure the sensitivities at a specific number of false positives per image. The FROC curve is an extension of the receiver operating characteristic (ROC) curve, which can be effective when multiple candidates are present in a single image. It plots the Recall (Sensitivity) against the average number of false positives per images. FROC is more sensitive at detecting small differences between performances and has higher statistical discriminative power [2].

4.1 Quantitative results

Figures 4(a) and (b) show the precision-recall curves corresponding to cilia detection for the CNN classifier applied after thresholding the template matching at different NCC levels (0.2, 0.3, 0.4, and 0.5), as well as the detection when using only template matching (which includes NCC thresholding at 0.546), as proposed in [15], for the training and test image, respectively. In the figures, the AUC is also stated. The results show that adding a CNN classifier significantly improves the AUC to 0.82 and 0.71 compared to the AUC of 0.48 and 0.42, for both the training and test image, respectively, at an NCC threshold level of 0.5.

The FROC curve for the proposed CNN applied to the training and test images when the template matching result was thresholded at different NCC levels (0.2, 0.3, 0.4, and 0.5) is shown in Fig. 5(a)–(b). This corresponds to the sensitivity of the classifier against total number of FP per image.

A classification confusion matrix is also shown in Table 2. The matrix shows the performance of the classifier for both the training and test image, in terms of TP (true positive), FP (false positive), FN (false negative), and TN (true negative), at equal error rate. At an NCC threshold level 0.5, the template matching method detected 212 (73 cilia and 139 non-cilia) candidates as potential cilia candidates. Amongst these, in the Table 2(A), the proposed CNN classifier correctly

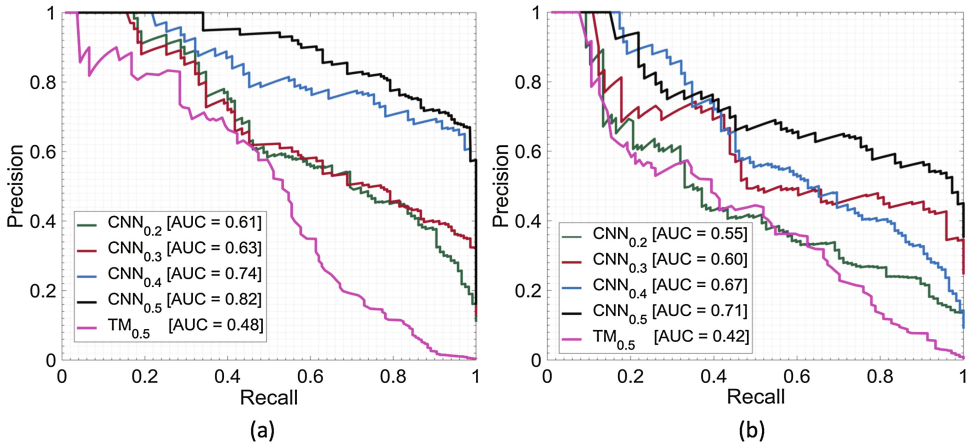


Fig. 4. Precision-recall curves of the CNN classifier at different NCC threshold levels shown together with the AUC for the template matching approach(TM) [15] for (a) training, (b) test images

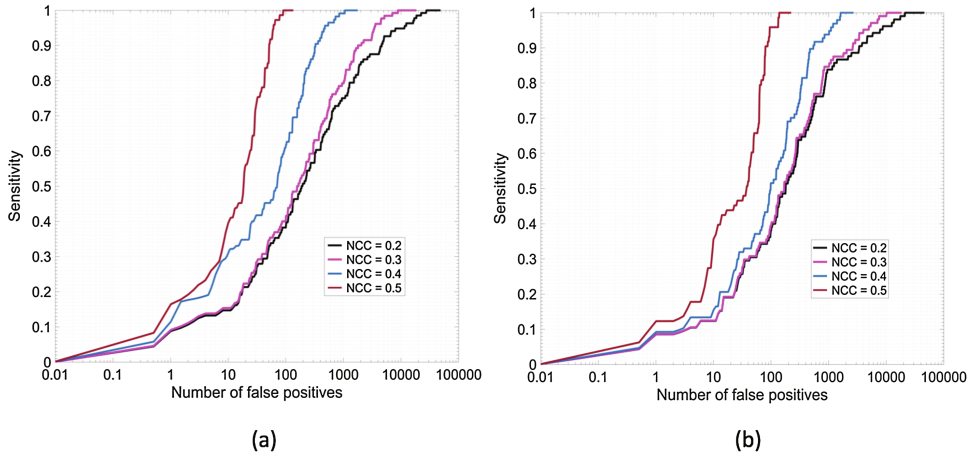


Fig. 5. FROC curves of the CNN classifier for (a) training image (b) test image at different NCC threshold levels. The number of FP are shown on a logarithmic scale.

classified 47 (TP) out of 73 (TP+FN) cilia candidates whereas from the set of 139 (FP+TN) non-cilia candidates, 26 non-cilia candidates (FP) were wrongly classified as cilia candidates by our proposed CNN classifier. We observe, in the training image, at equal error rate (Table 2(A)), the classifier also performed well when tested with the candidates extracted at an NCC threshold level of 0.4, but it eventually underperformed for the test image. The achieved results led us to finally conclude that the proposed CNN model yields a stable performance if it is incorporated with the candidates extracted at an NCC threshold level of 0.5. This observation is supported by the F-Score curves, shown in Fig. 6. Comparatively for

Table 2. Classification matrix of the CNN classifier at different NCC threshold levels for: (A) training image and (B) test image; at equal error rate.

A: Training image (Equal error rate)									
		0.2		**0.3**		**0.4**		**0.5**	
TP	FP	51	85	50	80	64	51	47	26
FN	TN	85	48 004	80	18 035	51	1 113	26	113
B: Test image (Equal error rate)									
		0.2		**0.3**		**0.4**		**0.5**	
TP	FP	38	67	37	66	37	60	37	36
FN	TN	67	45 926	66	18 348	60	2 658	36	188

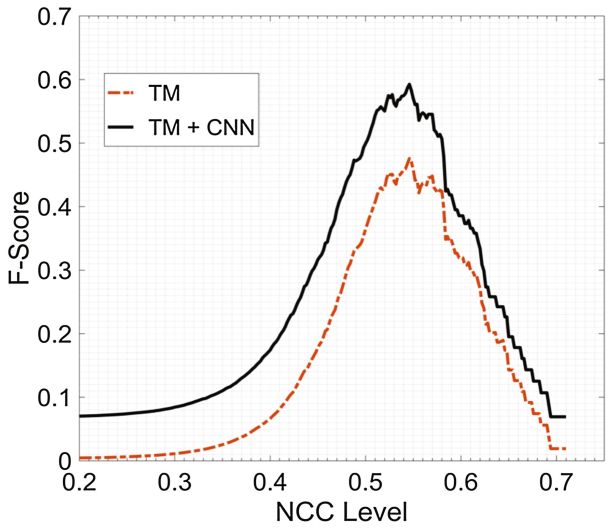


Fig. 6. F-score curves, for the test image, showing the improvement in overall performance by adding a CNN classifier with template matching approach(TM) [15] at different NCC threshold levels

the test image, at an NCC level of 0.546 (as suggested in [15]), the proposed CNN model increases the overall F-Score from 0.47 to 0.59.

4.2 False positive reduction results

Detection results of the proposed CNN model on a ROI of 650×650 for the test LM TEM image, at an NCC level of 0.5, are shown in Fig. 7(c)–(d). Figure 7(c) shows the detection results of the initial candidate detection step (template matching method, [15]) whereas Fig. 7(d) shows the improved results achieved by incorporating the proposed CNN model as an FP reduction step. In these images, the blue circles, red crossed circles, and green squares represent the candidates that have been correctly detected (TP), the candidates that have been

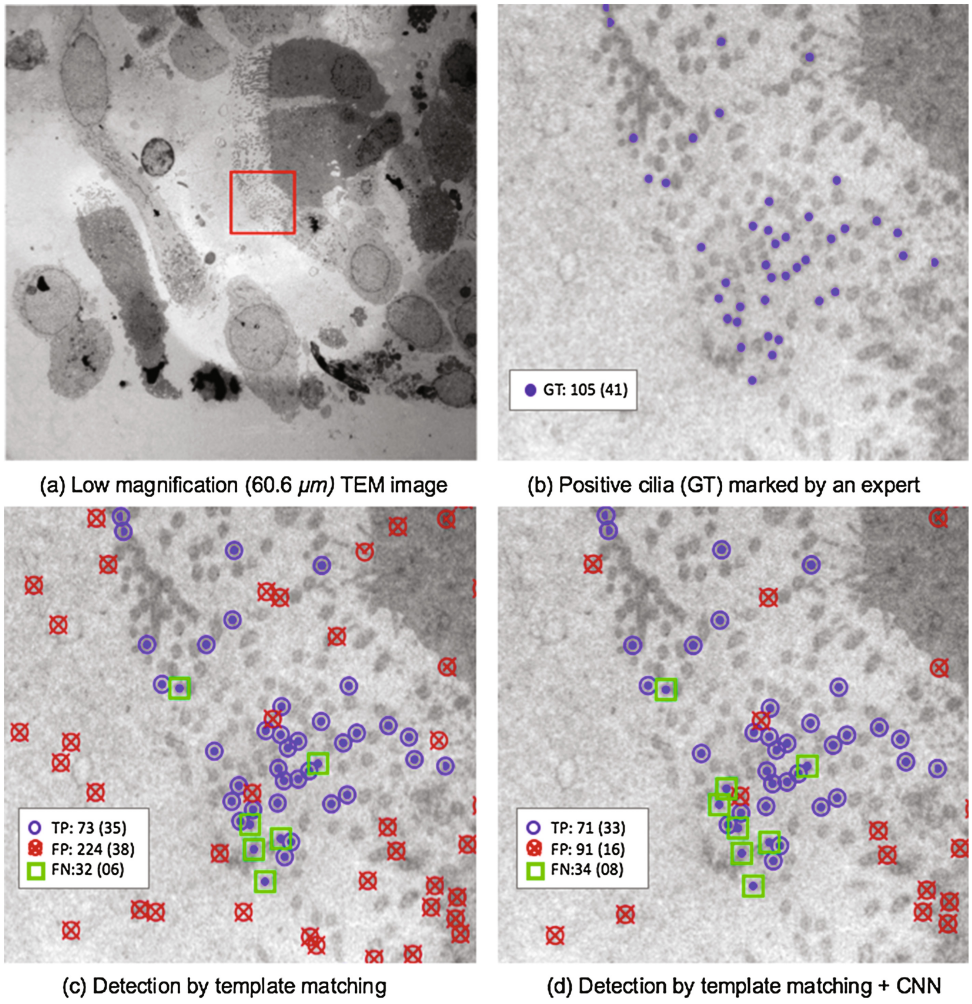


Fig. 7. Illustration of cilia detection results. (a) The 4096×4096 test image, (b) a 650×650 example subregion of the test image, (c) same subregion after initial template matching method, and (d) after proposed CNN classifier. The numbers are given for the whole image and for the ROI is in parenthesis. Here, blue circles, red crossed circles, and green squares represent the TP, FP, and FN, respectively. (Color figure online)

erroneously detected as cilia (FP), and the cilia that were missed with respect to the manually ascertained ground truth delineations and initial detection step (FN), respectively. These results show the potential of our CNN model for cilia detection in low magnification TEM images.

Examples of classified candidate image patches in the test image are shown in Fig. 8. The images marked in the first row are the TP and FP candidates from both methods (i.e., TM and CNN). In the second row, TP candidates detected by TM but erroneously classified as FN by CNN; and FP candidates detected by TM, which are successively classified as TN by proposed classifier.

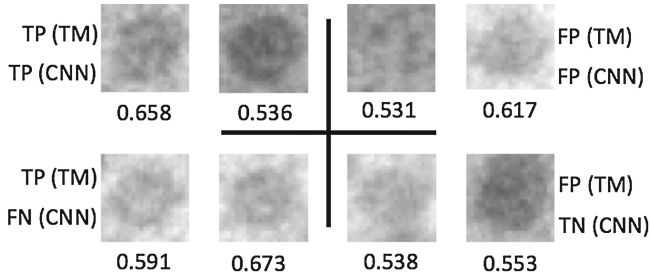


Fig. 8. Examples of candidates (with their corresponding NCC values) detected or missed by the proposed CNN model in the test image at an NCC level of 0.5. The first row shows TP's and FP's of both methods. The second row shows TP and FP candidates which are missed and successively classified by the CNN method, respectively.

5 Conclusion

In this paper, a CNN classifier is presented as a false positive reduction step for automated detection of cilia candidates in low magnification TEM images. The results suggest that adding a CNN classifier as a FP reduction step certainly improves the performance and results in an increased F-Score from 0.47 to 0.59. It was also investigated whether utilizing a CNN classifier as an additional refinement step would allow for using a lower NCC threshold in order to not discard true cilia objects in the template matching step. This was however, not found to be practically suitable as lowering the NCC threshold increases the number of candidates to analyze tremendously while only rather few additional true candidates are detected. It will be interesting in the future to develop and investigate a CNN model for the whole automated cilia detection problem, without relying on a first template matching step. This is currently not possible as it requires more training and test data.

Acknowledgments. The work is supported by Skype IT Academy Stipend Program, EU Institutional grant IUT19-11 of Estonian Research Council and the Swedish Innovation Agency's MedTech4Health program grant no. 2016-02329. J. Lindblad and N. Sladoje are supported by the Ministry of Education, Science, and Technological Development of the Republic of Serbia through projects ON174008 and III44006.

References

1. Brosch, T., Yoo, Y., Li, D.K.B., Traboulsee, A., Tam, R.: Modeling the variability in brain morphology and lesion distribution in multiple sclerosis by deep learning. In: Golland, P., Hata, N., Barillot, C., Hornegger, J., Howe, R. (eds.) MICCAI 2014. LNCS, vol. 8674, pp. 462–469. Springer, Cham (2014). doi:10.1007/978-3-319-10470-6_58
2. Chakraborty, D.: A status report on free-response analysis. Radiat. Prot. dosimetry **139**, 20–25 (2010)

3. Chatfield, K., Simonyan, K., Vedaldi, A., Zisserman, A.: Return of the devil in the details: delving deep into convolutional nets. In: British Machine Vision Conference (BMVC) (2014)
4. Chollet, F.: Keras (2015). https://github.com/fchollet/keras
5. Ciresan, D., Giusti, A., Gambardella, L.M., Schmidhuber, J.: Deep neural networks segment neuronal membranes in electron microscopy images. In: Advances in Neural Information Processing Systems, pp. 2843–2851 (2012)
6. Glorot, X., Bengio, Y.: Understanding the difficulty of training deep feedforward neural networks. In: Aistats, vol. 9, pp. 249–256 (2010)
7. Kingma, D., Ba, J.: Adam: a method for stochastic optimization. In: Proceedings of the 3rd International Conference on Learning Representations (ICLR) (2015)
8. Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. In: Advances in Neural Information Processing Systems, pp. 1097–1105 (2012)
9. LeCun, Y., Bottou, L., Bengio, Y., Haffner, P.: Gradient-based learning applied to document recognition. Proc. IEEE **86**(11), 2278–2324 (1998)
10. LeCun, Y., Hinton, G.: Deep learning. Nature **521**(7553), 436–444 (2015)
11. Li, R., Zhang, W., Suk, H.-I., Wang, L., Li, J., Shen, D., Ji, S.: Deep learning based imaging data completion for improved brain disease diagnosis. In: Golland, P., Hata, N., Barillot, C., Hornegger, J., Howe, R. (eds.) MICCAI 2014. LNCS, vol. 8675, pp. 305–312. Springer, Cham (2014). doi:10.1007/978-3-319-10443-0_39
12. Nair, V., Hinton, G.E.: Rectified linear units improve restricted Boltzmann machines. In: 27th International Conference on Machine Learning, pp. 807–814 (2010)
13. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. In: Proceedings of the 3rd International Conference on Learning Representations (ICLR) (2015)
14. Srivastava, N., Hinton, G.E., Krizhevsky, A., Sutskever, I., Salakhutdinov, R.: Dropout: a simple way to prevent neural networks from overfitting. J. Mach. Learn. Res. **15**(1), 1929–1958 (2014)
15. Suveer, A., Sladoje, N., Lindblad, J., Dragomir, A., Sintorn, I.M.: Automated detection of cilia in low magnification transmission electron microscopy images using template matching. In: 13th IEEE International Symposium on Biomedical Imaging (ISBI), pp. 386–390. IEEE (2016)
16. Szegedy, C., Liu, W., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., Rabinovich, A.: Going deeper with convolutions. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1–9 (2015)
17. Tieleman, T., Hinton, G.: Lecture 6.5-RmsProp: divide the gradient by a running average of its recent magnitude. COURSERA: Neural Networks for ML (2012)
18. Vogl, T.P., Rigler, A., Zink, W., Alkon, D.: Accelerating the convergence of the back-propagation method. Biol. Cybern. **59**(4–5), 257–263 (1988)
19. Xu, J., Xiang, L., Liu, Q., Gilmore, H., Wu, J., Tang, J., Madabhushi, A.: Stacked saparse autoencoder (SSAE) for nuclei detection on breast cancer histopathology images. IEEE Trans. Med. Imag. **35**(1), 119–130 (2016)

Publication B

Appeared in:

Bajić B*, Suveer A*, **Gupta A***, Pepic I, Lindblad J, Sladoje N, and Sintorn I-M.

Denoising of Short Exposure Transmission Electron Microscopy Images for Ultrastructural Enhancement. Accepted for publication In *Proceedings of the 15th International Symposium on Biomedical Imaging (ISBI)*, Washington D.C., USA, April 2018.

*These authors have contributed equally.

DENOISING OF SHORT EXPOSURE TRANSMISSION ELECTRON MICROSCOPY IMAGES FOR ULTRASTRUCTURAL ENHANCEMENT

Buda Bajić[1*], *Amit Suveer*[2*], *Anindya Gupta*[3*], *Ivana Pepić*[1]
Joakim Lindblad[2,4], *Nataša Sladoje*[2,4], *Ida-Maria Sintorn*[2,5]

[1] Faculty of Technical Sciences, University of Novi Sad, Serbia
[2] Centre for Image Analysis, Dept. of Information Technology, Uppsala University, Sweden
[3] T. J. Seebeck Dept. of Electronics, Tallinn University of Technology, Estonia
[4] Mathematical Institute of the Serbian Academy of Sciences and Arts, Belgrade, Serbia
[5] Vironova AB, Stockholm, Sweden

ABSTRACT

Transmission Electron Microscopy (TEM) is commonly used for structural analysis at the nm scale in material and biological sciences. Fast acquisition and low dose are desired to minimize the influence of external factors on the acquisition as well as the interaction of electrons with the sample. However, the resulting images are very noisy, which affects both manual and automated analysis. We present a comparative study of block matching, wavelet domain, energy minimization, and deep convolutional neural network based approaches to denoise short exposure high-resolution TEM images of cilia. In addition, we evaluate the effect of denoising before or after registering multiple short exposure images of the same imaging field to further enhance fine details.

Index Terms— Denoising, Convolutional Neural Networks, TEM, Cilia

1. INTRODUCTION

Transmission Electron Microscopy (TEM) is an imaging technique providing nm resolution. It is therefore well suited and often used to analyze structural details in biological samples and tissue sections for research and clinical diagnostics. However, both manual and automated analysis of TEM images are negatively affected by a number of imaging factors, such as sample preparation artefacts, non-optimal microscope alignment and focusing, electrons interacting with and modifying the sample, and motion artefacts from e.g. sample drift and vibrations. Preprocessing with an aim to enhance the relevant details (ultrastructures) is often applied.

The imaging artefacts can be reduced by decreasing the electron dose and acquisition time. However, this results in images with more noise and increases the need for denoising. The noise induced by TEM is non-additive and signal-dependent. It can be modeled by a mixed Poisson-Gaussian (PG) distribution [1, 2]. However, in short exposure images, the Gaussian noise dominates. We consider three classical methods suited for Gaussian and PG noise: a block matching [3], wavelet domain [4], and energy minimization [5] based method, and evaluate their performances on short exposure TEM images of cilia in a cell section sample, Fig. 1. Moreover, observing that convolutional neural networks (CNNs) have recently been shown to perform well in denoising [6, 7], we have developed a suitable denoising CNN model and included it in the comparison. To the best of our knowledge, this is the first denoising CNN evaluated on TEM noisy data.

Denoising is commonly performed on a single image. However, our ultimate goal is to enhance fine details in TEM images, which, in theory, can be achieved by generating synthetic long exposure images by aggregating (median) a number of short exposure ones. We, therefore, also investigate two strategies of combining aggregation and denoising: (1) co-registration and aggregation of a number of short exposure images is performed first and followed by denoising of the aggregation; (2) short exposure images are denoised and the resulting ones are then co-registered and aggregated. Enhancement of structural information by registration and aggregation of scanned lines, images or objects, is commonly used in other biomedical imaging techniques, e.g., in scanning transmission electron microscopy (STEM) [2, 8], and cryo-EM [9].

2. DENOISING METHODS

2.1. Block-matching and 3D filtering (BM3D)

Block matching based techniques utilize self similarities present in the image. The BM3D algorithm [3] is suitable for images with structural redundancy, which is common in biological images, and also in our case. BM3D has successfully been used for denoising light microscopy images [10] and STEM images [11].

*These authors have contributed equally.

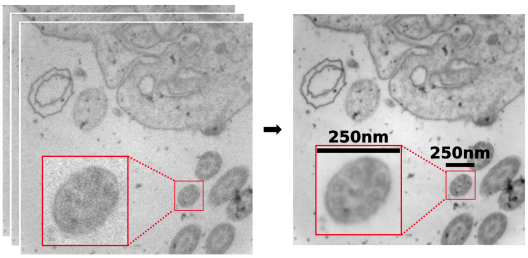


Fig. 1: **Left**: Short exposure TEM image (2048×2048 pixels) from a series of 100 images. **Right**: Ground truth created by co-registration and aggregation of the stack to the left. The two insets show magnified views (250×250 pixels) of one cilium.

2.2. PURE-LET

In the PURE-LET [4] method the denoising process is expressed as a linear expansion of thresholds (LET). The threshold optimization solely relies on a data-adaptive unbiased estimation of the mean squared error (MSE), derived in a non-Bayesian framework (PURE: Poisson–Gaussian unbiased risk estimate, defined in the Haar wavelet domain). The method is suitable for light microscopy images, as presented in the original paper, and it also performs well in restoring STEM images as shown in [12].

2.3. Energy minimization (EM)

Many denoising methods are based on solving an inverse problem through energy minimization. We perform denoising by minimizing an energy function which includes a quadratic data fidelity term, suited for Gaussian noise, and a regularization term which provides numerical stability and noise suppression. We use Total Variation (TV) regularization [13] smoothed by the Huber potential function [14], resulting in well preserved edges in images [5]. We have previously shown applicability of this approach to cilia ultrastructure enhancement in long exposure images [15], where we applied a generalized version of the method suited to PG noise and blind deblurring.

2.4. Denoising Convolutional Neural Network (DCNN)

Inspired by the good performance of the approaches in [16, 17], we jointly train two CNNs as an ensemble. The architecture is shown in Fig. 2. The training of both streams is performed on image patches of size 128×128 with an overlapping stride of 16 pixels. Prior to the training, the patches are normalized to the range [0,1]. The first stream consists of four convolution blocks, two transposed convolution blocks and one residual block. The convolution block encodes the image representations while removing the noise, whereas the

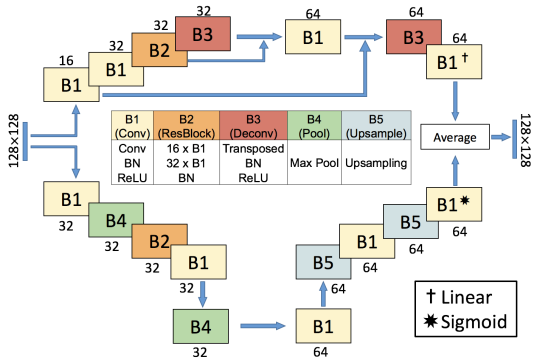


Fig. 2: The two-stream DCNN architecture. The sizes of output feature maps of each block are shown on top of each block and generated using 3×3 convolutions. The last 1×1 convolution blocks of each stream use linear and sigmoid activation, respectively, instead of ReLU.

transposed convolution block decodes these representations to restore the noise-free image content. The residual block contains two convolution blocks. Batch normalization (BN) [18] is used as regularization before rectified linear unit (ReLU) activation to deal with internal covariate shift. To elevate the training performance, skip connections are used and followed by a BN layer. During experiments, we found that the prediction made by the first stream restores most content with blur. Considering that, we incorporated a second stream consisting of four convolution blocks, two up-sampling blocks, two max-pooling layers, and one residual block. The reconstructed output of the second stream contains high-frequency content, however, with an inconsistent illumination in respect to corresponding ground truth. Motivated by the above observations, we performed an end-to-end training by averaging the outputs of both streams, which resulted in an improved output. We used stochastic gradient descent (SGD) to optimize the weights in a mini-batch scheme of 16 patches. The initial learning rate was set to 0.001, and reduced to 1/10 of the current value after every epoch. We used MSE and binary cross-entropy as loss function. The DCNN is implemented using Tensorflow backend in Keras [19] and trained for 15 epochs in a five-fold cross validation scheme. The average training time is 300 s/epoch on a GPU GeForce GTX 1080.

3. EXPERIMENTS AND RESULTS

3.1. Quantitative evaluation

The dataset consists of a series of 100 noisy short exposure (2 ms) images, captured at the same spatial location in the cell section sample (FoV=2000 nm). All images are of size 2048×2048 pixels and acquired with the low-voltage

Table 1: Results on the test data set. Average PSNR and SSIM (± SD) over 90 single images are given in the 1st and 2nd rows. Rows 3 and 4 contain average PSNR and SSIM over 18 aggregated groups of 5 short exposure images followed by denoising. Average PSNR and SSIM over 18 images each obtained by aggregating 5 denoised short exposure images, are given in rows 5 and 6. Optimal parameters (in parentheses) estimated during the training phase are used. Best performances are bolded.

		Initial	BM3D (σ_{bm})	PURE-LET (σ_{pl})	EM (λ)	DCNN
Denosing of single images	PSNR	22.25	37.39 ± 0.30 (105)	38.44 ± 1.09 (75)	37.80 ± 0.27 (0.25)	**38.04** ± 0.21
	SSIM	0.019	0.233 ± 0.007 (95)	0.219 ± 0.007 (55)	**0.255** ± 0.027 (0.20)	0.252 ± 0.002
Denoising of 5 aggregated noisy images	PSNR	27.88	40.45 ± 1.09 (95)	40.19 ± 1.06 (35)	40.19 ± 0.54 (0.125)	**40.86** ± 0.37
	SSIM	0.037	0.270 ± 0.019 (35)	0.263 ± 0.017 (25)	0.277 ± 0.017 (0.10)	**0.282** ± 0.011
Aggregation of 5 denoised single images	PSNR	22.25	39.65 ± 1.04 (95)	40.21 ± 0.48 (55)	39.92 ± 0.93 (0.10)	**40.84** ± 0.45
	SSIM	0.019	0.261 ± 0.013 (25)	0.265 ± 0.011 (45)	0.273 ± 0.021 (0.075)	**0.276** ± 0.009

MiniTEM[1] system. A low-noise image, used as a ground-truth, is estimated by registering each short exposure image to the first image of the series using rigid registration, followed by aggregating the information by computing the pixel-wise median value, illustrated in Fig. 1.

We utilize 10 images for the training of the DCNN[2] and explorative parameter tuning of each method – the regularization weight λ for EM, and the expected std. of Gaussian noise, σ_{bm} and σ_{pl}, for BM3D[3] and PURE-LET[4], respectively. The tuned parameters are used to compare the performance of each method on the remaining 90 images. Apart from evaluating the performances on denoising single images, we additionally tune parameters and evaluate the performances of the methods when used to 1) denoise the resulting image after registering and aggregating (median) five short exposure images, and 2) when registering and aggregating (median) five denoised short exposure images.

The performance is evaluated using well known and often used the peak-signal-to-noise ratio (PSNR) and structural similarity index measure (SSIM) [20]. As indicated in [21], different levels of degradations applied to the same image can yield the same PSNR. We observe that PSNR performs poorly on discriminating structural content in images which plays an important role in ultrastructural analysis of TEM images. As SSIM is proposed with the aim to compare structural changes in images imitating what the human visual system does, this measure is considered a more reliable measure of visual similarity of images.

Denoising of single short exposure images – The average PSNR and SSIM over all 90 images from the test dataset are given in Table 1, along with the parameters tuned during the training. The EM method marginally outperforms the remaining methods in terms of SSIM. On the other hand, DCNN outperforms all classical methods in terms of PSNR. A cilium from a single noisy image and the corresponding denoised instances obtained with all 4 methods are presented in the first row of Fig. 3.

Denoising of 5 aggregated short exposure images – We register groups of 5 short exposure images and aggregate them by the pixel-wise median. We denoise the resulting 18 images by all 4 considered methods. The average PSNR and SSIM (over 18 images) are given in Table 1. As confirmed by both average PSNR and SSIM, the DCNN method outperforms the other methods. A noisy cilium instance from aggregating 5 short exposure images and the corresponding denoised results obtained with all 4 methods are presented in the middle row of Fig 3.

Aggregation of 5 denoised short exposure images – We denoise 5 sequentially acquired short exposure images, then register them and aggregate them by the pixel-wise median. The average PSNR and SSIM for the 18 resulting images are given in Table 1. The corresponding results on the cilium subimage are shown in the bottom row of Fig. 3. Note that the first image is the ground truth, i.e., the median aggregated 100 short exposure images. In this strategy as well, the DCNN produces the highest PSNR and SSIM.

3.2. Qualitative evaluation

To validate the level of agreement between the quantitative results and visual (qualitative) results, we performed a subjective visual evaluation conducting a two-step voting process by six experts. In the first step, involving only the classical methods, the experts rated the results (1st, 2nd, 3rd best) on the cilium subimage produced by each of the methods with different parameter settings. The displayed images (7 for each method) spanned a parameter range centered around the maximal SSIM for that method. The procedure was repeated for the two strategies of aggregating 5 short exposure images (denoising prior or post registering and aggregation). The second step involves all four methods. The images resulting from the two aggregation strategies utilizing the tuned parameter settings as decided in Step 1, together with the DCNN results were displayed (random, unknown order) and the experts rated them again (as the 1st, 2nd, 3rd best). The denoised image with the majority of votes is highlighted in Fig. 3.

[1] Vironova AB, Stockholm, Sweden

[2] https://bitbucket.org/anindya_gupta/tem-denoising/

[3] http://www.cs.tut.fi/~foi/GCF-BM3D

[4] http://bigwww.epfl.ch/algorithms/denoise/

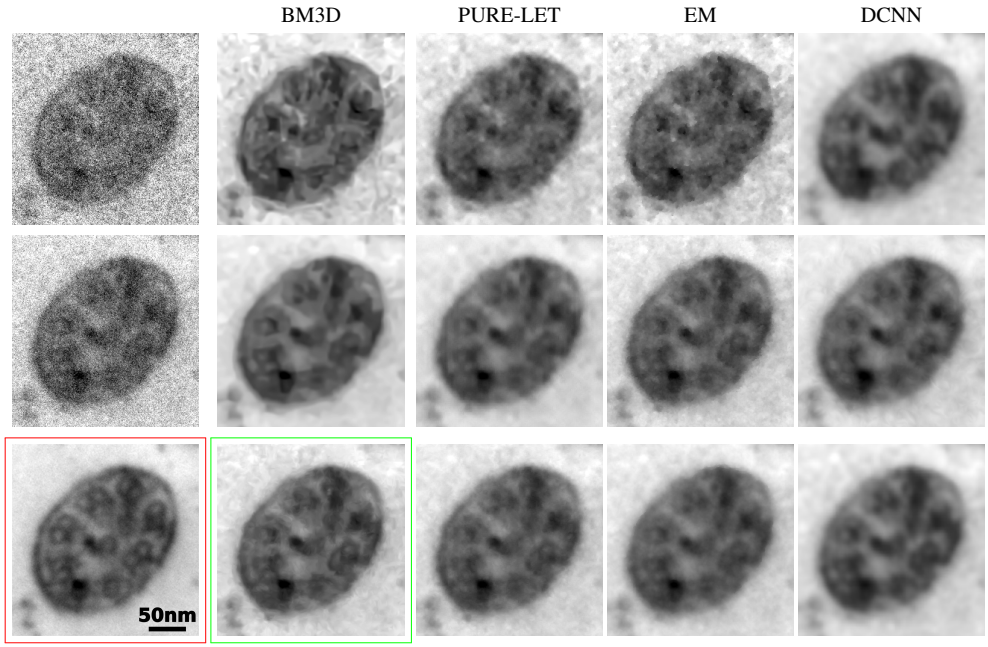


Fig. 3: Noisy and denoised close ups of a cilium instance obtained with the considered methods. **Top**: Denoising of a single image. **Middle**: Denoising of 5 aggregated noisy images. **Bottom**: Aggregation of 5 denoised single images. The red frame (bottom left) indicates the ground truth. The green frame indicates the best ranked image in the two-step visual voting process.

In the first step of the voting procedure, the experts' votes agreed well with the quantitative results based on SSIM. However, in some cases, the experts visually preferred slightly less regularized images. This is not surprising since humans prefer to see sharp details and can "ignore" noise to some degree. The results corresponding to maximal PSNR were consistently judged as over-regularized.

4. DISCUSSION AND CONCLUSION

Short exposure time reduces the influence of motion blur and electron interaction with the sample. That, however, affects the image quality. We have quantitatively and qualitatively compared four different denoising methods that can be used to improve the resulting poor image quality. To additionally enhance ultrastructural information in TEM images, we have investigated two strategies i.e., denoising of aggregated series of noisy images and aggregation of several denoised short exposure images of the same view.

From the quantitative and qualitative results in Table 1 and Fig. 3 it is clear that denoising can improve both single and multiple aggregated short exposure images. Comparatively, noisy single images require more regularization. It is also interesting to note that the optimal parameter values for the classical methods differ a lot depending on whether single short exposure or aggregated images are to be denoised. Note that the DCNN was only trained on single frames also for the strategies using aggregated images. Overall, DCNN gives the highest quantitative scores, but based on the visual assessment BM3D applied to noisy images prior to aggregation produced the most appealing result.

Both of the two aggregation strategies, denoising the registered and aggregated image or registering and aggregating after denoising the short exposure images, improve the results approximately equally well. One advantage with the former aggregation strategy is that only one denoising computations is performed instead of five.

5. ACKNOWLEDGMENT

This work is supported by VINNOVA, MedTech4Health grant 2016-02329, the Ministry of Education, Science, and Techn. Development of the Rep. of Serbia (proj. ON174008 and III44006), Swedish Research Council grant 2014-4231, the IT Acad. Prog. and EU Inst. grant IUT19-11 of ERC.

6. REFERENCES

[1] M. Vulović, E. Franken, R. B. G. Ravelli, L. J. van Vliet, and B. Rieger, "Precise and unbiased estimation of astigmatism and defocus in Transmission Electron Microscopy," *Ultramicroscopy*, vol. 116, pp. 115–134, 2012.

[2] B. Berkels and B. Wirth, "Joint denoising and distortion correction of atomic scale Scanning Transmission Electron Microscopy images," *arXiv preprint arXiv:1612.08170*, 2016.

[3] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising by sparse 3-D transform-domain collaborative filtering," *IEEE Trans. Image Process.*, vol. 16, no. 8, pp. 2080–2095, 2007.

[4] F. Luisier, T. Blu, and M. Unser, "Image denoising in mixed Poisson–Gaussian noise," *IEEE Trans. Image Process.*, vol. 20, no. 3, pp. 696–708, 2011.

[5] B. Bajić, J. Lindblad, and N. Sladoje, "Restoration of images degraded by signal-dependent noise based on energy minimization: an empirical study," *J. of Electronic Imaging*, vol. 25, no. 4, pp. 043020–043020, 2016.

[6] X. Mao, C. Shen, and Y-B. Yang, "Image restoration using very deep convolutional encoder-decoder networks with symmetric skip connections," in *Advances in Neural Info. Process. Systems 29*, pp. 2802–2810. 2016.

[7] J. Kim, J. Kwon Lee, and K. Mu Lee, "Accurate image super-resolution using very deep convolutional networks," in *IEEE Conf. on Comp. Vis. and Patt. Recog. (CVPR)*, 2016.

[8] A. B. Yankovich, C. Zhang, A. Oh, T. J. A. Slater, F. Azough, R. Freer, S. J. Haighand, R. Willett, and P. M. Voyles, "Non-rigid registration and non-local principle component analysis to improve electron microscopy spectrum images," *Nanotechnology*, vol. 27, no. 36, pp. 364001, 2016.

[9] Y. Cheng, N. Grigorieff, P. A. Penczek, and T. Walz, "A primer to single-particle cryo-electron microscopy," *Cell*, vol. 161, no. 3, pp. 438–449, 2015.

[10] J. Boulanger, C. Kervrann, P. Bouthemy, P. Elbau, J. B. Sibarita, and J. Salamero, "Patch-based nonlocal functional for denoising fluorescence microscopy image sequences," *IEEE Trans. on Med. Imaging*, vol. 29, no. 2, pp. 442–454, 2010.

[11] N. Mevenkamp, P. Binev, W. Dahmen, P. M. Voyles, A. B. Yankovich, and B. Berkels, "Poisson noise removal from high-resolution STEM images based on periodic block matching," *Adv. Structural and Chemical Imaging*, vol. 1, no. 1, pp. 3, 2015.

[12] N. Mevenkamp, A. B. Yankovich, P. M. Voyles, and B. Berkels, "Non-local means for Scanning Transmission Electron Microscopy images and Poisson noise based on adaptive periodic similarity search and patch regularization," in *Vision Modeling and Visualization*, 2014, pp. 63–70.

[13] L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear Total Variation based noise removal algorithms," *Physica D: Nonlinear Phenomena*, vol. 60, pp. 259–268, 1992.

[14] R. Schultz and R. Stevenson, "Stochastic modeling and estimation of multispectral image data," *IEEE Trans. Image Process.*, vol. 4, no. 8, pp. 1109–1119, 1995.

[15] B. Bajić, J. Lindblad, and N. Sladoje, "Blind restoration of images degraded with mixed Poisson-Gaussian noise with application in Transmission Electron Microscopy," in *IEEE Int. Symp. on Biomed. Img. (ISBI 2016), proceedings*, 2016, pp. 123–127.

[16] C. Ledig, L. Theis, F. Huszar, J. Caballero, A.P. Aitken, A. Tejani, J. Totz, Z. Wang, and W. Shi, "Photo-realistic single image super-resolution using a generative adversarial network," in *IEEE Conf. on Comp. Vis. and Patt. Recog. (CVPR)*, 2017.

[17] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with Deep Convolutional Neural Networks," in *Advances in Neural Info. Proc. Systems 25*, pp. 1097–1105. 2012.

[18] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *Int. Conf. on ML*, 2015, pp. 448–456.

[19] F. Chollet, "Keras," https://github.com/fchollet/keras, 2015.

[20] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, 2004.

[21] Z. Wang and A. C. Bovik, "Mean squared error: Love it or leave it? A new look at signal fidelity measures," *IEEE Signal Process. Magazine*, vol. 26, no. 1, pp. 98–117, 2009.

Publication C

Appeared in:

Gupta A, Saar T, Märtens O, and Moullec YL.

Automatic detection of multi–size pulmonary nodules in CT images: Large-scale validation of false-positive reduction step. *Medical Physics*, 2018:45(3):1135-49. doi: 10.1002/mp.12746

Publication D

Submitted to:

Gupta A, Saar T, Märtens O, Moullec YL, and Sintorn I-M.

Detection of Pulmonary Micronodules in CT Images and False Positive Reduction Using 3D Convolutional Neural Networks. *IEEE Journal of Biomedical and Health Informatics*

Detection of Pulmonary Micronodules in CT Images and False Positive Reduction Using 3D Convolutional Neural Networks

Anindya Gupta, Tõnis Saar, Olev Märtens, *Member, IEEE,* Yannick Le Moullec, and Ida-Maria Sintorn

***Abstract*—Purpose: Micronodules are small lesions that are radiologically characterized to manifest the fatal and incurable occupational pulmonary disease silicosis. Identifying scattered micronodules in computed tomography (CT) scans is a tedious and challenging task for radiologists. We present a novel CAD system specifically dedicated to detect micronodules in thoracic CT scans.**

Method: The proposed system consists of a candidate-screening module and a false positive reduction module. The candidate-screening module is initiated by a lung segmentation algorithm which is refined using a morphological closing operation to include small lesions attached to pleura. Next, the micronodules are identified through a combination of 2D/3D rule based thresholding and morphological operation steps. In the false positive (FP) reduction module, each identified candidate is classified by a 3-D convolutional neural networks (CNN). It automatically encodes the discriminative representations derived from training data by exploiting the volumetric information of each candidate.

Result: On 598 CT scans of the publically available LIDC/IDRI database, the CAD system achieves detection sensitivities of 74.3% (648/872) and 86.7% (756/872) at 1 FP/scan and 8 FPs/scan, respectively.

Conclusion: Our proposed CAD system efficiently identifies micronodules in thoracic scans with only a small number of FPs. Our experimental results showed that the automatically generated features by the 3-D CNN are highly discriminant, making it a well-suited FP reduction module of a CAD system.

***Index Terms*—Computed tomography, convolutional neural networks, false positive reduction, micronodules.**

I. Introduction

OCCUPATIONAL pulmonary diseases are one of the most common causes of lung impairments around the world. Amongst all, silicosis is one of the prevalent and incurable abnormalities following long and continuous exposure (more than 5 years) to silica dust. In current radiological practices, silicosis is typically characterized as widespread, well-defined solid pulmonary micronodules < 4 mm [12], [17]. Existence of silicosis is well known since ancient times; but it has in present times grown to a global public health problem, reporting thousands of cases every year. Although, the prevalence has decreased compared with past decades, it is still an obscured and underreported disease [6]. Late stage silicosis can lead to lung cancer [22], thus early detection is essential to control the progression.

The current guideline of the International Labor Organization (ILO) recommends to employ chest radiography as the primary diagnostic modality for the progression analysis of silicosis. Since imaging and clinical history are critical diagnostic resources for silicosis progression analysis, clinicians are, in addition, required to perform a follow-up examination relating the radiological manifestation and dust exposure history [10]. Radiographs can visualize micronodules in quite advanced stages, but the micronodules are visible and detectable in much earlier stages by computed tomography (CT) due to its volumetric characterization and high sensitivity, thus enabling the early manifestation of small nodules [16].

Aiming at early manifestation of pulmonary nodules, lung cancer screening trials have already suggested CT as the primary imaging modality. The outcome of the National Lung Screening Trial (NLST) shows that CT is better and more accurate than conventional radiographs, reporting a substantial amount of micronodules in the reference repository, i.e., the Lung Image Database Consortium (LIDC) and Image Database Resource Initiative (IDRI) [1]. Such high sensitivity, however, realizes at a cost of enormous amounts of volumetric data, resulting in increased reading efforts for radiologists during routine practices and screening trials, which can typically take up to 15 min/scan [16]. Manual detection and marking of small lesions, specifically micronodules, is still challenging due to absence of symptoms and high similarities to cross-sectional vessels (as shown in Fig. 1). It is thus essential to develop computer-assisted detection (CAD) schemes to facilitate clinicians in this monotonous, error-prone and time-consuming process.

II. Background

CAD schemes typically consist of two modules i.e., candidate screening and false positive reduction. In the candidate screening module, methods comprising of multiple operations (e.g., intensity thresholding, contextual and morphological operations) aim to detect a considerable amount of candidates while rapidly screening through the CT scan, resulting in high sensitivity at a high false positive (FP) rate. The FP reduction module aims at reducing the high FP rate using a set of discriminative features within an empirically optimized classifier [7]. In such a sequential setting, methods in the

In part, this project has received funding from the European Unions Horizon 2020 research and innovation programme under grant No 668995. This work is also supported by Estonian IT Academy Stipend Program, EU Institutional grant IUT19-11 of Estonian Research Council and European Regional Development Fund, North Estonia Medical Center, Doctoral School of ICT.

A. Gupta, O. Martens, and Y. Le Moullec are with the Thomas Johann Seebeck Dept. of Electronics, Tallinn University of Technology, Tallinn, 12611 Estonia. (anindya.gupta@ttu.ee).

T. Saar is with the Eliko Technological Center Tallinn, 12611 Estonia.

I-M Sintorn is with the Center of Image Analysis, Dept. of Information Technlogy, Uppsala University, Uppsala, Sweden. (ida.sintorn@it.uu.se).

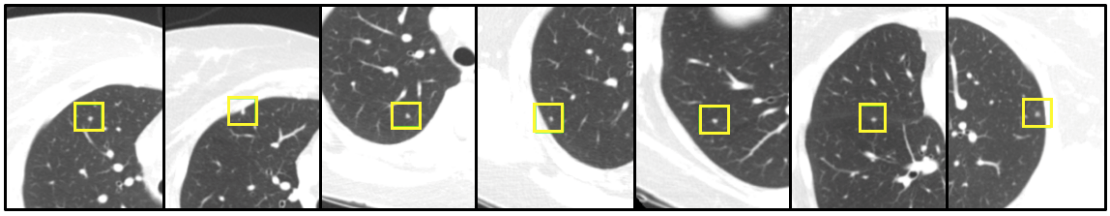


Fig. 1: Seven axial CT slices illustrating the detection task for micronodules (< 4 mm). The micronodules are shown in the yellow bounding boxes. The manual detection of micronodules for silicosis is indeed a tedious and challenging task.

candidate-screening module are often unspecific while the FP reduction module indeed stands as the decisive module and performs an in-depth assessment on candidates to provide the final detection results.

The FP reduction module usually relies on an underlying set of extracted intensity, contextual and texture features. Generally, extraction of discriminative features is in high correlation with the quality of the results from the candidate-screening module (i.e., precise segmentation and enhancement of potential candidates). In addition, such conventional or handcrafted features also tend to suffer from limited representation capability and are insufficient to deal with the large contextual variation [4], which could negatively affect the performance of the classifier. Lately, several methods have reported promising results by employing deep learning techniques, specifically, convolutional neural networks (CNNs) in the FP reduction module. For instance, Roth *et al.* [19] proposed a 2.5D CNN model for lymph node detection using image patches from three axes-oriented 2D patches of a 3D volume. Setio *et al.* [20] presented a fusion of several 2D multi-view CNNs to learn the discriminative representations from different axes-oriented 2D patches of nodules (3-30 mm) and reported a state-of-the-art sensitivity of 90.1% at 4 FP/scan on the benchmark LIDC/IDRI dataset. Although their method demonstrated the effectiveness of the CNN, it was still limited in the full utilization of the 3D spatial information. Lately, Dou *et al.* [7] presented a multilevel contextual 3D CNN framework by integrating three different sizes of receptive fields and surpassed previous works employing handcrafted features or 2D CNN with a reported sensitivity of 90.7% at 4 FP/scan.

Despite the progress in the development of CAD schemes for automatic detection of pulmonary nodules, very little work has been reported for the automatic detection of micronodules in CT scans. Initially, Brown *et al.* [2] presented a method for the detection of micronodules using 15 CT scans containing a total of 57 micronodules. They employed thresholding based segmentation of lunch parenchyma and segmented micronodule candidate by filtering on anatomical features (size, shape and location). No further FP reduction step was applied, resulting in a sensitivity of 70% with 15 FP/scan. Following that, Zhao *et al.* [25], proposed a CAD scheme for the detection of small nodules in low dose CT. They utilized a local density maximum (LDM) method to identify the potential candidates and incorporated 2D/3D features for further FP reduction. The method was validated on 28 multi-slice CT scans including 165 nodules, resulting in an overall sensitivity of 66.7% with 8.7 FPs/scan. Lately, Jacobs *et al.* [16] presented the first automated CAD scheme for early detection of micronodules by employing a lung segmentation, template matching, and a k-nearest neighbor classifier. The system was validated on 54 low-dose CT scans from a controlled study group of a private lung cancer screening trial and yielded a sensitivity of 84% with an average of 8.4 FPs/scan.

Although these results are encouraging, it is still possible to boost the performance by exploiting the discriminative capabilities of deep neural networks for the automatic detection of micronodules. This will have several advantages over the conventional methods. For instance, these techniques directly learn discriminative features and effectively leverage the feature interaction and hierarchy within itself. In addition, the performance of network-crafted features are improved in a systematic fashion within the optimization of the same model.

A. Contributions

In this study, we present an automated CAD system for the detection of micronodules in thoracic CT scans. The system is built and evaluated on the LIDC/IDRI dataset. The major focus of this work is to develop an efficient FP reduction module, which can also be integrated with other candidate detection methods to further improve the performance of CAD schemes for micronodules in general. We demonstrate the potency of encoding discriminative representations from the 3D CNN in complicated anatomical surrounding environments to improve the detection accuracy. In addition,this is, to the best of our knowledge, first study to present an automated system for the detection of micronodules evaluated on the LIDC/IDRI dataset.

III. Material

We used CT scans from the LIDC/IDRI dataset to train and validate our method. The LIDC/IDRI is the largest publicly accessible dataset of annotated thoracic CT scans [1]. It consists of 1018 CT scans captured at seven institutions with heterogeneous image acquisition and reconstruction parameters. Four expert radiologists annotated every case in a two-phase process to maintain the inter-reader variability. During the first (blind) annotation process, each radiologist independently annotated the suspicious lesions as either non-nodule, nodule < 3 mm, or nodule ≥ 3 mm. In the second process, each radiologist

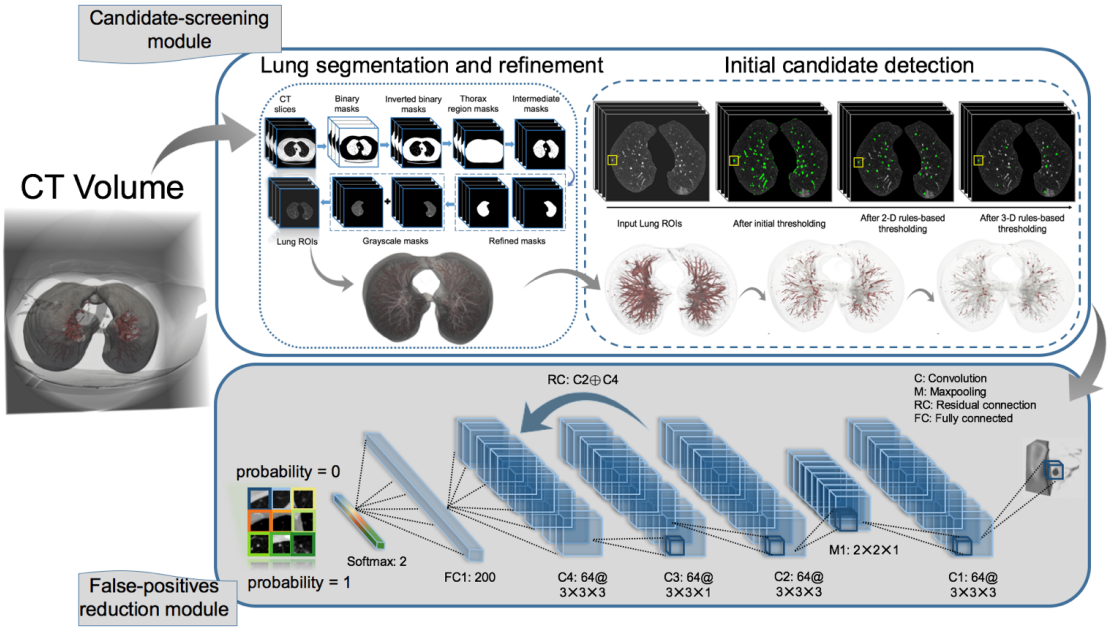


Fig. 2: An overview of the proposed CAD system. The system is divided into a candidate-screening and a false positive reduction module. Initial candidates are detected from the segmented lung ROIs using 2D and 3D features-based thresholding operations. The false positive reduction module is implemented using a 3D CNN. The architecture of the proposed 3D CNN is shown using an example of an extracted 3D scans of $20\times20\times7$ voxels with the candidate in the center. The locally connected convolutional layers encode highly discriminative representations for a given candidate. Once the discriminative representations are computed, the candidates are connected to the classifier (softmax) via a fully-connected (FC) layer. The Softmax performs classification by predicting the final probability for each candidate in the range of 0 to 1. The grey arrows represent the general flow of the CAD.

compared their marking with the markings from the other radiologists to conclude the final decision.

The annotations include the subjective ratings and outer boundary markings of nodules ≥ 3 mm and only the center-of-mass for nodules < 3 mm due to their smaller size and challenging interpretation. We selected CT scans with slice thickness below 3 mm for training and validation purposes, resulting in 598 CT scans. Thicker CT scans were rejected due to their unsuitability in current clinical practices [24]. The distribution of section thickness (z-) across the 598 CT scans taken from the LIDC-IDRI dataset are shown in Fig. 3. The pixel sizes range from 0.461 to 0.977 mm in the x- and y-dimensions. We considered a volume of 34 mm^3 corresponding to a sphere size of 4 mm in diameter as a size threshold criterion and selected only those lesions that were agreed on by at least two radiologists, resulting in 872 micronodules in the reference set.

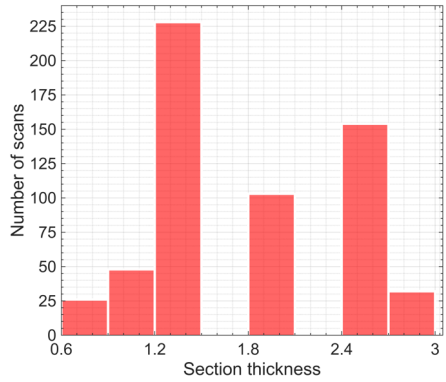


Fig. 3: Distribution analysis of slice thickness across the 598 CT scans taken from the LIDC/IDRI dataset.

IV. Method

The outline of the automated CAD scheme for detection of micronodules is shown in Fig. 2. It consists of a candidate-screening module with lung segmentation and candidate detection steps, and a FP reduction module with candidate extraction and a 3D CNN to classify the micronodule candidates.

A. Candidate screening module

1) Preprocessing: We resampled the CT scans to an isotropic voxel size of 0.6 mm^3 using cubic interpolation to

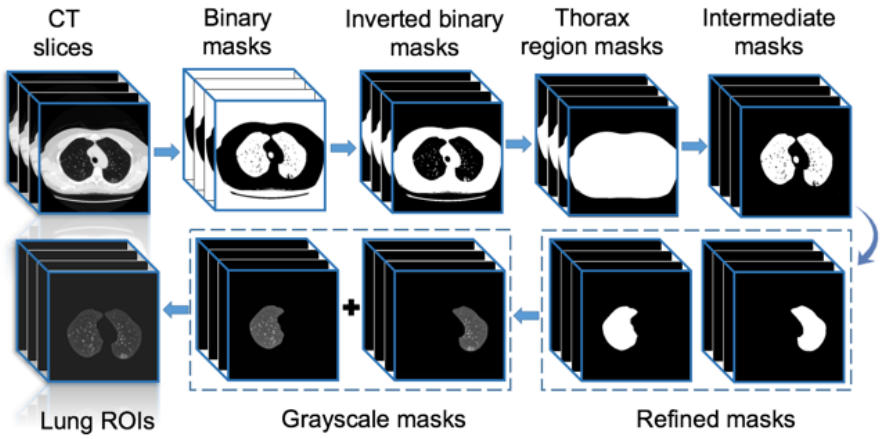


Fig. 4: Overview of the lung segmentation process based on the extraction of multiple masks.

deal with the inconsistencies across different CT scans in the candidate-screening module. In addition, having isotropic resolution is often beneficial for the implementation of image analysis operations.

2) Lung Segmentation: The lung regions are extracted by the method reported in [11], and the steps are illustrated in Fig. 4. The method proceeds section-by-section and first applies a threshold of -400 HU to create a *binary mask* for each slice. Next, the *binary mask* is inverted and morphological hole-filling and 2D connected component labeling is applied to extract the largest component as the *thorax region mask*. An *intermediate mask* consisting of two lung lobes and the trachea is identified by the logical AND applied to the *thorax region mask* and *thorax region mask*. Next, the trachea is removed by identifying the two largest components (left and right lungs) by applying morphological hole-filling, followed by 2D connected component labeling and area based filtering. Note that, if the area of the second largest component is less than half of the area of the largest component, the single component is considered as a lung. The two identified components are later superimposed on two *black background mask* of the same dimensions as the input image. To refine these *lung mask*, a morphological closing using a disk shaped structuring element with 1% (5 pixels) size of the original image is applied. Next, the *grayscale masks* are extracted by superimposing the *refined masks* on the input image. Finally, the lung ROI is obtained by adding the individual *grayscale masks*.

3) Initial candidate detection: Localization of micronodules is a challenging task due to their high similarity with cross-sectioned vessels. We observed during multiple experiments that a high threshold diminishes micronodules whereas lower thresholds introduce a large number of false positives. We hence, apply an empirically computed threshold of -700 HU on the extracted lung regions, followed by a morphological erosion using a disk shaped structuring element of a 1-pixel radius on the labeled objects. Subsequently, we employ a sub-algorithm module consisting of 2D and 3D features-based thresholding to identify the initial candidates. We utilized area (*R1*) and eccentricity (*R2*) as 2D rules, resulting in high sensitivity with a large number of false positives. Area is calculated as the total number of pixels in the object region times the pixel size in mm^3, whereas eccentricity (ranges from zero to 1) is defined as the ratio of the length of the major axis and the length of the minor axis of the object. An object will be considered for further processing if the following expression is true: $(\mathrm{R1} > 15 \,\&\&\, \mathrm{R1} < 78) \,\&\&\, \mathrm{R2} > 0.70)$ where && corresponds to the logical AND. This allows to distinguish between a circular object and a sticklike object in the 2D slices. The 2D features can possibly eliminate some of the connected voxels at the beginning or end of a (true) 3D object (micronodule), but they are still beneficial in order to exclude large and sparsely connected 3D (false) objects. Prior to 3D features-based thresholding, 3D connected component labelling using 26-point connectivity scheme is performed.

Next, we apply elongation and sphericity as 3D feature thresholds on the candidates of size up to $34\,mm^3$ (equivalent to a 4 mm diameter sphere) to further eliminate FPs. Elongation (*R3*) is defined as the ratio of the minimum dimension in the $x, y,$ or z direction over the largest dimension in any direction. Whenever a candidate satisfies the criterion i.e., $\mathrm{R3} > 0.60$, it is considered for the next rule. The sphericity (*R4*) is defined as $6\sqrt{\pi}VA^{-3/2}$, where V and A are the volume and area of an object. Only those that fulfills the criterion i.e., $\mathrm{R4} > 0.40$, will be considered as plausible small candidates. These defined parameters and conditions ensure that the considered objects have relatively compact shapes in three dimensions (*R3* and *R4*) and fall into the desired size range. The specific selection of rules and their thresholds were empirically determined in a pilot study (consisting of 25% of randomly chosen CT cases from the training set). The results of the candidate detection sub-algorithms, applied on an example slice, are shown in Fig. 5.

B. False positive reduction module

1) Candidate extraction: Small lesions in anisotropic CT scans typically range over up to nine voxels (x, y) and four

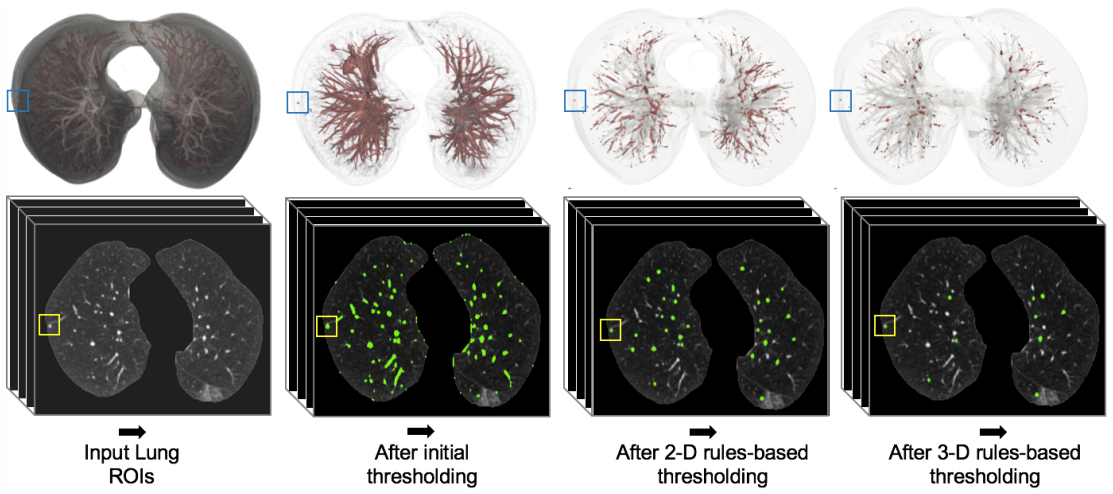


Fig. 5: Outputs of the features-based candidate detection module. The first row shows the flow on a 3D volume, wherein candidates are marked in red and the true micronodule is placed in the blue bounding box. The second row shows the flow in 2D, when applied on an example slice of the same volume wherein plausible candidates are marked in green and the true micronodule is placed in the yellow bounding box. The border of all images are trimmed for illustration purposes.

slices [8] due to the lower sampling and, hence, resolution in the z direction. We extracted scans of 20×20×7 voxels from the original anisotropic CT scans, to include sufficient contextual information, as input to the classifier. For example, if a candidate is centered at $V_{(x,y,z)}$, we extract an input volume $\hat{v} = V_{(x-(m+1):x+m,y-(m+1):y+m,z-k:z+k)}$, where V is the CT volume, m=10, and k=3.

2) 3D Convolutional Neural Network architecture: The architecture of our proposed 3D CNN is as shown in the bottom half of Fig. 2. The network architecture and its hyper-parameters, such as number of layers, kernel size, and learning rate, were determined during multiple experiments. We cascade four convolutional blocks, one pooling layer, one fully connected layer, and a softmax layer to encode the discriminative volumetric representations (feature maps). The convolutional block consists of one 3D convolution layer followed by batch normalization as regularization and rectified linear unit (ReLU) nonlinear activation. Each 3D convolutional layer (consisting of multiple neuron activated feature maps) encodes diverse representations by convolving 3D kernels over the output of the preceding layer. All convolutional blocks generate 64 feature maps using 3×3×3 convolutions, except the third convolutional layer which generates 64 maps using 3×3×1 convolutions to simplify the feature concatenation step. We used a maximum pooling layer with a pool size of 2×2×1 to downsample the output of the first convolutional block. The maximum pooling layer considers the maximum activation in non-overlapping cubic windows to encode the translation and scale invariant representations. The fully-connected (dense) layer, consisting of 200 neurons, is densely connected to the flattened output from the preceding layer and is followed by a softmax layer to yield the final probability distribution predicted by the network.

Generally, variations in the parameters from each layer (i.e., internal covariate shift) slow down the network training due to saturating nonlinearities and requires a low learning rate. This can adversely affect the training of the CNN with a risk of poor generalization performance [15]. Lately, batch normalization (BN) has enabled the CNNs to learn faster with better generalization of the network and overcome the issue of internal covariate shift. While training with BN, each feature map computed by a linear operation (here convolution) is normalized separately over the mini-batch to have a mean (μ) of zero and variance (σ^2) of 1. For example, a layer with an input $\mathbb{X} = (x_1,...,x_m)$, where m is the total number of feature maps computed after applying a linear operation (here, 64 feature maps after the 1^{st} convolutional layer). Each x_n is formed by all the corresponding feature maps of the candidates in the mini-batch (here, 128). The BN for n^{th} feature map can be expressed as:

$$\hat{x}_n = \frac{x_n - \mu(x_n)}{\sqrt{\sigma^2 [x_n]}} \quad (1)$$

However, just simply normalizing the feature map can constrain the representation capabilities of the network. Therefore, a pair of learning parameters (learned along with the original model parameters) for scaling by γ_n and shift by β_n is applied to the normalized feature map $\hat{x}_n$ as:

$$y_n = \gamma_n \hat{x}_n + \beta_n$$

We applied BN between each layer of 3D convolutions and nonlinear activation to reduce the internal covariate shift and to accelerate the network training. For an additional performance gain, we employed a residual (skip) connection by adding the output of the second convolution block to the input of the fourth convolutional blocks. Residual connections have lately

shown an improvement in the classification performance [13]. The skip connections are connections, which skip one or more layers and perform identical mapping by summing the input of one layer to the output of at least one skipped layer. The residual mapping is based on the approximation of the residual function instead of the original one directly from a 3D convolutional layer $\mathcal{H}(\cdot)$, and is expressed as:

$$\mathcal{H}(x_{out}) = x_{in} + \mathcal{F}(x_{in}, \{W_k\})$$

where, x_{in} and x_{out} are its input and output; $\mathcal{F}(\cdot)$ is a 3D residual mapping associated with a set of parameters $\{W_k\}$ where k=1, i.e., skipping 1 convolutional block, consisting of a 3D convolutional layer, a BN layer and ReLU nonlinear activation layer. From a learned feature weights sharing perspective, a residual connection enables feature reuse at no extra parameters and computational complexity. In addition, the gradient can easily flow through skip connections during the backward pass of the training.

3) Candidate augmentation: To deal with the severe class imbalance (here, 1 true candidate for every 447 false candidate), we performed augmentation of the existing true candidates. The issue of training with unbalanced datasets is that the learned representations can be skewed towards the most frequent sample type, resulting in limited or biased capabilities of the trained classifier. We used translation, flipping, and rotation to perform the augmentation. First, we randomly translated the centroid of the candidate by ± 2 voxels, then flipped in left-right and up-down directions, followed by rotation by [90°, 180°, 270°], resulting in a set of altogether 0.25 million true candidates. By this, we preserve the anatomical surrounding of the candidates and allow the classifier to encode rotation and translation invariant representations. Prior to the classification, we clipped the intensities of each candidate to the interval (-1000, 1000 HU) and normalized to the range (0, 1).

4) Training: On the given training dataset, RMSProp [23] was used to efficiently optimize the weights of the 3D CNN. RMSProp is an adaptive optimization algorithm, which normalizes the gradients by utilizing the magnitude of recent gradients. The weights were initialized using normalized initialization as proposed in [9] and updated in a mini-batch scheme of 128 candidates (as described above). The biases were initialized with zero and the learning rate was set to 0.001. A dropout [21] of 0.5 was implemented on the output of the dense layer to avoid overfitting. Softmax loss (cross-entropy error loss) was utilized to measure the error loss. With a set of N training sample pairs $\{X_i, Y_i\}_{i=1...,N}$ and Θ parameters of the network, where Y_i is the label corresponding to the input sample X_i, the cross-entropy loss ($\mathcal{L}$) is computed as:

$$\mathcal{L} = -\frac{1}{N}\sum_{i=1}^{N} Y_i \cdot \log p(Y_i|X_i;\Theta)$$

The 3D CNN model was implemented using Tensorflow backend in Keras [5] and trained for 50 epochs. The average training time was approximately 350 s/epoch on a GPU GeForce GTX 1080.

C. Evaluation strategy

To evaluate the candidate-screening module, we employed a criterion that if a candidate lies within a three-pixel range of the centerofmass of the respective micronodule from the reference set then it is considered as a true positive candidate, else a FP candidate.

To evaluate the FP reduction module, we used a five-fold cross-validation scheme, and measured two performance metrics: 1) Area under the ROC curve (AUC) and 2) Competition Performance Metric (CPM) [18]. The AUC shows the performance of CNN on classifying candidates. The CPM measures the average sensitivity at different operating points of the free-response operating characteristic (FROC) curve [3]: 1/8, 1/4, 1/2, 1, 2, 4, and 8 FPs/scan. The FROC curve plots the Recall (Sensitivity) against the average number of FPs per scan. It is more sensitive at detecting small differences between performances when multiple lesions are present in a single scan and has higher statistical discriminative power. We also computed the 95% confidence interval using bootstrapping with 1 000 bootstraps, as detailed in [8], to determine the upper and lower confidences of the CAD system.

V. Result and Discussion

We first report on a systematic evaluation of crucial CNN architecture parameters and then report the overall performance of our proposed CAD system.

A. Analysis of network configuration

To investigate performance with regard to the network configuration, we conducted a series of experiments to analyze the impact of three crucial parameters: 1) Batch normalization (BN) layer, 2) residual (skip) connection, and 3) size of receptive field in the z- direction. The architecture and parameters for analyzing the effect of the skip connection and receptive field is shown in Table I.

TABLE I: Architecture and parameters of CNN configurations investigated. Here, each convolution (C) layer employs 64 filters and M1, RC and FC1 layers stands for maxpooling, residual connection and fully-connected, respectively.

	3-D CNN$_{Residual}$		3-D CNN		3-D CNN$_{Elongated_{RF}}$	
Layer	Kernel	Parameters	Kernel	Parameters	Kernel	Parameters
C1	(3×3×3)	2K	(3×3×3)	2K	(5×5×5)	808
M1	(2×2×1)	N/A	(2×2×1)	N/A	(2×2×2)	N/A
C2	(3×3×3)	110K	(3×3×3)	110K	(3×3×3)	110K
C3	(3×3×1)	37K	(3×3×1)	37K	(3×3×3)	110K
RC	Yes	N/A	No	N/A	Yes	N/A
C4	(3×3×3)	110K	(3×3×3)	110K	(3×3×3)	110K
FC1	200	320K	200	320K	250	1024K
Softmax	2	402	2	402	2	502

1) Effect of batch normalization layers: To evaluate the impact of batch normalization (BN) layers on the network, we repeated the experiments without incorporating it in the 3D CNN with residual connection. Figure 6 shows the comparison in terms of error loss and accuracy across the number of epochs on the validation set. As observed, the BN improves the overall performance of the classifier through faster convergence. In addition to the speed improvements, it also

enables the use of higher learning rates while overcoming the problem of saturated nonlinearities. Overall, the batch normalized model achieves higher validation accuracy, which is due to its regularizing effect and more stable gradient propagation. It is, thus, worthwhile to incorporate the BN layers in the network since it prevents model divergence and results in better generalization.

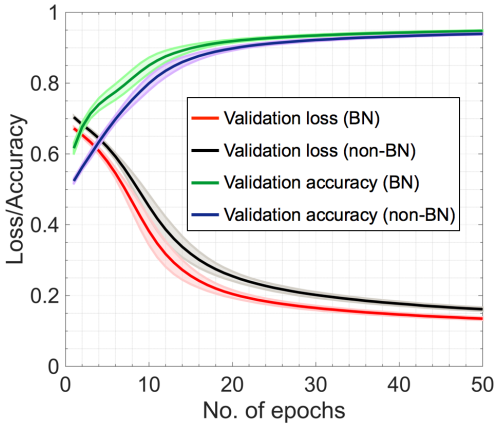


Fig. 6: Comparison of the mean accuracy and mean loss for the networks with and without batch normalization (BN) layer on the five-folds of cross-validation sets. The shaded area spans the maximum and the minimum scores for each plot.

2) Effect of residual (skip) connections: The inclusion of a residual (skip) connection for summing-up the learned representations (features) is worthwhile to overcome the gradient degradation and also to improve the overall performance of the CAD system. To investigate this, we repeated the experiments without incorporating a skip connection in the same network architecture (the middle architecture in Table I compared to the left architecture). Table II shows that including the skip connection certainly alleviates the overall performance with a CPM score of 0.727. We consent that the representations of the 3D CNN employing skip connection exhibit better discriminability in comparison to the one without it.

3) Size of receptive field: Inspired by the good performance of the method [14] on pulmonary nodule detection, we also conducted experiments using an elongated volume ($20\times20\times20$) to train a 3D CNN. We intended to observe the influence of the size of receptive field in the z- direction on the detection performance of the classifier, since the input images have different slice thicknesses. The receptive field of a network is the size of the volume, which can influence the prediction in a position. As can be noticed from Fig. 7, the 3D CNN with the elongated receptive field as input was not able to yield a substantial performance in comparison to the network with an input of ($20\times20\times7$). This is because the large receptive field possesses more redundancy due to much larger contextual surrounding in training, perhaps more generalized to ambiguous contextual information, resulting in degraded discrimination capabilities of the network. As reported in [8], we also believe that the amount of surrounding contextual information exploited by the network has a great impact on the final predicted probabilities of the classifier. It is, thus, crucial yet challenging to determine an optimal receptive field for the detection of micronodules. Although we have shown a preliminary comparison, determining an optimal receptive field is still an interesting and open topic for future research.

B. Comparison of conventional and CNN features

For comparative purposes, we computed a set of 27 conventional features, typically used for developing traditional pulmonary CAD systems; to train a shallow three-layer artificial neural networks (ANN) model. We computed 21 intensity and 6 morphological features from two spherical regions centered on the candidate with a diameter of 4 mm and 10 mm, respectively. The first region considers the intensity in a spherical region around the segmented candidate whereas the second region takes into consideration the intensity of a wider contextual surrounding of the lesions. The set of intensity features consist of energy, entropy, skewness, variance, kurtosis, maximum, minimum, mean, standard deviation of both regions, and the ratio of maximum, minimum and mean intensities of the first region to the second region. The six morphological features consist of the size dimensions (mm) of the segmented candidates in the $x, y,$ and z directions, volume, eccentricity, and sphericity. All features were normalized to zero mean and unit standard deviation. We used $L2$ regularization to control the model overfitting. The weights were optimized using scaled conjugate gradient descent (SGD) algorithm in 3 000 iterations. Softmax loss function (cross-entropy error loss) was used to measure the loss.

Table II shows that the conventional features were not able to yield a very high CPM compared to the CNNs features. This is because the conventional features are highly dependent on the results from the candidate-screening module. The conventional features are typically affected by the similar intensity distribution of nearby blood vessels and tissues in the segmentation results, and thus resulting in less discriminating power. In addition, task-specific feature engineering is a challenging and time-consuming task, and even optimized feature sets often result in uncertainties when testing on independent heterogeneous datasets. Considering the sensitivity to subtle changes and complicated optimization process of traditional feature extraction techniques, it is highly beneficial to employ the 3-D CNNs to extract discriminating features. Although our reported observations exhibit the potency of neuron-crafted features over the conventional features, careful optimization of CNNs is still a crucial requisite to obtain a substantial performance throughput.

C. Quantitative and Qualitative Analysis

On the full dataset of 598 CT scans, the 2D rules of the candidate-screening module detected 94.8% (827/872) micronodules at an average 670 FPs/scan. These false positives were further reduced by the 3D rules, where after 91.6% (799/872) of all micronodules at an average of 447 candidates/scan were detected. For quantitative comparison, the FROC curves for the different 3D CNN configurations tested on the dataset are

TABLE II: Quantitative performance of the CAD system with different CNN configurations on the LIDC/IDRI dataset. Sensitivities at 7 operative points along with CPM and AUC are also listed.

FPs/scan	1/8	1/4	1/2	1	2	4	8	CPM	AUC(A_z)
3-D CNN$_{residual}$	0.549	0.629	0.680	0.743	0.792	0.832	0.867	0.727	0.988
3-D CNN	0.505	0.571	0.642	0.707	0.772	0.793	0.845	0.691	0.957
3-D CNN$_{Elongated}$	0.280	0.401	0.485	0.593	0.682	0.760	0.814	0.573	0.943
Shallow NN	0.125	0.330	0.377	0.441	0.510	0.556	0.609	0.421	0.884

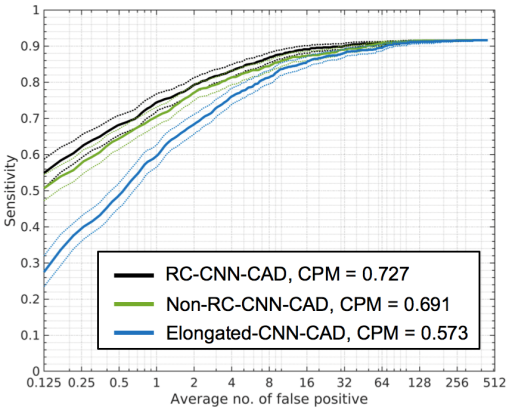


Fig. 7: FROC curve for the different CNN configurations tested on the LIDC/DIRI dataset. Here, the RC (residual connection) and Non-RC configurations are trained with an input of 20×20×7 and the Elongated configuration is trained with an input of 20×20×20 elongated receptive field. The dashed curve shows the 95% bootstrap confidence interval. The number of FPs are shown on a logarithmic scale.

shown in Fig. 7. The average number of FPs are shown on a logarithmic scale. The figure also shows the average CPM and the 95% bootstrap confidence interval.

For each network evaluated as the FP reduction module, the sensitivities at seven operating points along with average CPM and AUC are provided in Table II. Given a set of candidates for the classification task, the 3D CNNs reach an AUC of 0.988. By employing the best 3D CNN configuration, the proposed method identifies 74.3% (648/872) and 86.7% (756/872) of the micronodules at 1 FP/scan and 8 FPs/scan, respectively. Note that the maximum sensitivity of the classifier is bounded by the sensitivity of the initial candidate detection module, i.e., 91.6% (799/872). This indicates that the classification stage correctly classifies 81.1% (648/799) and 94.6% (756/799) of the initially detected candidates at 1 and 8 FPs/scan, respectively.

Some examples of detected micronodules and false positives are shown in Fig. 8. We observed that a substantial number of false positives detected at 1 FP/scan are small vessels, nodular- like structures, and scaring. All these structures depict the same characteristics as micronodules and manual interpretation of these small lesions can be exhaustive and challenging. Unlike the radiologists detection of micronodules, vessels that have nodule like appearances in the plane should not mislead the computer interpretation. This is because the detection algorithm can distinguish between a spherical small lesion and a cylindrical vascular structure.

The FP reduction module was validated in a five-fold cross validation scheme. However, the CT cases from the LIDC/IDRI were still minimally used to optimize the candidate-screening module. For the completeness of the proposed system, further testing on an independent set is still an essential requisite. Since our FP reduction module can be employed independently to the proposed candidates-screening module, the overall sensitivity can possibly be alleviated by integrating with more efficient candidate-screening modules. In addition, the practical usefulness and significance of our method can be determined by performing an observer study with and without it. Although the dedicated CAD systems are developed with a primary aim of efficient diagnosis of CT scans, the proposed detection system can alternatively be employed as a second reader to further validate the primary annotations.

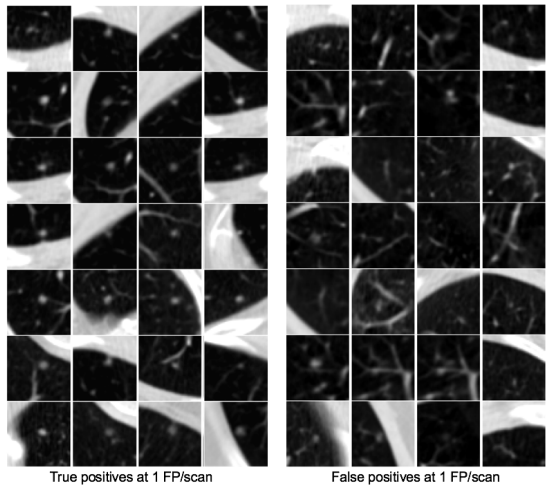


Fig. 8: Examples of lesions detected by the CAD system. The left set of lesions are micronodules detected at 1 FP/scan. The right set shows the false positives candidates detected at 1 FP/scan.

VI. Conclusion

Using 598 annotated CT scans from the LIDC/IDRI, we present an automated CAD system specifically designed to detect pulmonary micronodules with a diameter of up to 4 mm. The proposed system integrates a candidate-screening module

and a FP reduction module. A high detection sensitivity of the candidate-screening module is crucial to determine the upper-bound quality of the CAD system. Aiming at high sensitivity, we designed a simplified candidate-screening module which detects most of the candidates but at a relatively high FP rate. To further eliminate the large number of FPs, we presented a novel 3D CNN framework as an efficient FP reduction module. We showed that the proposed FP reduction module can detect the vast majority of highly suspicious lesions in thoracic CT scans at an expense of only small number of false positives. Hence, we conclude that our proposal CAD system employing 3D CNN as a FP reduction module could be highly beneficial for radiologists to identify small lesions and to overcome the labor intensive process of interpretation. The promising results shows that the proposed CAD system can be considered as an assistive tool.

REFERENCES

[1] Samuel G Armato, Geoffrey McLennan, Luc Bidaut, Michael F McNitt-Gray, Charles R Meyer, Anthony P Reeves, Binsheng Zhao, Denise R Aberle, Claudia I Henschke, Eric A Hoffman, et al. The lung image database consortium (lidc) and image database resource initiative (idri): a completed reference database of lung nodules on ct scans. *Medical physics*, 38(2):915–931, 2011.

[2] Matthew S Brown, Jonathan G Goldin, Robert D Suh, Michael F McNitt-Gray, James W Sayre, and Denise R Aberle. Lung micronodules: automated method for detection at thin-section ctinitial experience. *Radiology*, 226(1):256–262, 2003.

[3] DP Chakraborty. A status report on free-response analysis. *Radiation protection dosimetry*, 139(1-3):20–25, 2010.

[4] Jie-Zhi Cheng, Dong Ni, Yi-Hong Chou, Jing Qin, Chui-Mei Tiu, Yeun-Chung Chang, Chiun-Sheng Huang, Dinggang Shen, and Chung-Ming Chen. Computer-aided diagnosis with deep learning architecture: applications to breast lesions in us images and pulmonary nodules in ct scans. *Scientific reports*, 6:24454, 2016.

[5] François Chollet et al. Keras, 2015.

[6] Christian W Cox, Cecile S Rose, and David A Lynch. State of the art: imaging of occupational lung disease. *Radiology*, 270(3):681–696, 2014.

[7] Qi Dou, Hao Chen, Lequan Yu, Jing Qin, and Pheng-Ann Heng. Multilevel contextual 3-d cnns for false positive reduction in pulmonary nodule detection. *IEEE Transactions on Biomedical Engineering*, 64(7):1558–1567, 2017.

[8] Bradley Efron and Robert J Tibshirani. *An introduction to the bootstrap*. CRC press, 1994.

[9] Xavier Glorot and Yoshua Bengio. Understanding the difficulty of training deep feedforward neural networks. In *Proceedings of the thirteenth international conference on artificial intelligence and statistics*, pages 249–256, 2010.

[10] Michael I Greenberg, Javier Waksman, and John Curtis. Silicosis: a review. *Disease-a-Month*, 53(8):394–416, 2007.

[11] Anindya Gupta, Tonis Saar, Olev Martens, and Yannick Le Moullec. Automatic detection of multisize pulmonary nodules in ct images: Large-scale validation of the false-positive reduction step. *Medical physics*, 45(3):1135–1149, 2018.

[12] David M Hansell, Alexander A Bankier, Heber MacMahon, Theresa C McLoud, Nestor L Muller, and Jacques Remy. Fleischner society: glossary of terms for thoracic imaging. *Radiology*, 246(3):697–722, 2008.

[13] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

[14] Xiaojie Huang, Junjie Shan, and Vivek Vaidya. Lung nodule detection in ct using 3d convolutional neural networks. In *Biomedical Imaging (ISBI 2017), 2017 IEEE 14th International Symposium on*, pages 379–383. IEEE, 2017.

[15] Sergey Ioffe and Christian Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. *arXiv preprint arXiv:1502.03167*, 2015.

[16] C Jacobs, SHTT Opdam, EM van Rikxoort, OM Mets, J Rooyackers, PA de Jong, M Prokop, and B van Ginneken. Automated detection and quantification of micronodules in thoracic ct scans to identify subjects at risk for silicosis. In *Medical Imaging 2014: Computer-Aided Diagnosis*, volume 9035, page 90351I. International Society for Optics and Photonics, 2014.

[17] Kun-Il Kim, Chang Won Kim, Min Ki Lee, Kyung Soo Lee, Choong-Ki Park, Seok Jin Choi, and Jong Gi Kim. Imaging of occupational lung disease. *Radiographics*, 21(6):1371–1391, 2001.

[18] Meindert Niemeijer, Marco Loog, Michael David Abramoff, Max A Viergever, Mathias Prokop, and Bram van Ginneken. On combining computer-aided detection systems. *IEEE Transactions on Medical Imaging*, 30(2):215–223, 2011.

[19] Holger R Roth, Le Lu, Ari Seff, Kevin M Cherry, Joanne Hoffman, Shijun Wang, Jiamin Liu, Evrim Turkbey, and Ronald M Summers. A new 2.5 d representation for lymph node detection using random sets of deep convolutional neural network observations. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 520–527. Springer, 2014.

[20] Arnaud Arindra Adiyoso Setio, Francesco Ciompi, Geert Litjens, Paul Gerke, Colin Jacobs, Sarah J van Riel, Mathilde Marie Winkler Wille, Matiullah Naqibullah, Clara I Sánchez, and Bram van Ginneken. Pulmonary nodule detection in ct images: false positive reduction using multi-view convolutional networks. *IEEE transactions on medical imaging*, 35(5):1160–1169, 2016.

[21] Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. Dropout: A simple way to prevent neural networks from overfitting. *The Journal of Machine Learning Research*, 15(1):1929–1958, 2014.

[22] Kyle Steenland and Elizabeth Ward. Silica: a lung carcinogen. *CA: a cancer journal for clinicians*, 64(1):63–69, 2014.

[23] Tijmen Tieleman and Geoffrey Hinton. Lecture 6.5-rmsprop: Divide the gradient by a running average of its recent magnitude. *COURSERA: Neural networks for machine learning*, 4(2):26–31, 2012.

[24] Dong Ming Xu, Hester Gietema, Harry de Koning, René Vernhout, Kristiaan Nackaerts, Mathias Prokop, Carla Weenink, Jan-Willem Lammers, Harry Groen, Matthijs Oudkerk, et al. Nodule management protocol of the nelson randomised lung cancer screening trial. *Lung cancer*, 54(2):177–184, 2006.

[25] Binsheng Zhao, Gordon Gamsu, Michelle S Ginsberg, Li Jiang, and Lawrence H Schwartz. Automatic detection of small lung nodules on ct utilizing a local density maximum algorithm. *journal of applied clinical medical physics*, 4(3):248–260, 2003.

Publication E

Appeared in:

Lidayová K, **Gupta A**, Frimmel H, Sintorn I-M, Bengtsson E, Smedby Ö.

Classification of Cross-sections for Vascular Skeleton Extraction Using Convolutional Neural Networks. In *Proceedings of the 21th Medical Image Understanding and Analysis (MIUA)*, Edinburgh, Scotland, July 2017, CCIS-723 pp. 182-194. doi: 10.1007/978-3-319-60964-5_16

Classification of Cross-sections for Vascular Skeleton Extraction Using Convolutional Neural Networks

Kristína Lidayová[1], Anindya Gupta[2], Hans Frimmel[3], Ida-Maria Sintorn[1], Ewert Bengtsson[1], and Örjan Smedby[4]

[1] Centre for Image Analysis, Dept. of IT, Uppsala University, Uppsala, Sweden
kristina.lidayova@it.uu.se
[2] T. J. Seebeck Dept. of Electronics, Tallinn University of Technology, Estonia
[3] Division of Scientific Computing, Dept. of IT, Uppsala University, Sweden
[4] School of Tech. and Health, KTH Royal Institute of Technology, Stockholm, Sweden

Abstract. Recent advances in Computed Tomography Angiography provide high-resolution 3D images of the vessels. However, there is an inevitable requisite for automated and fast methods to process the increased amount of generated data. In this work, we propose a fast method for vascular skeleton extraction which can be combined with a segmentation algorithm to accelerate the vessel delineation. The algorithm detects central voxels - nodes - of potential vessel regions in the orthogonal CT slices and uses a convolutional neural network (CNN) to identify the true vessel nodes. The nodes are gradually linked together to generate an approximate vascular skeleton. The CNN classifier yields a precision of 0.81 and recall of 0.83 for the medium size vessels and produces a qualitatively evaluated enhanced representation of vascular skeletons.

Keywords: Vascular skeleton, CT angiography, Convolutional neural networks, Classification

1 Introduction

Vascular diseases are among the leading causes of death around the world. To diagnose a vascular disease, a detailed description of the state of each major artery in the arterial tree is needed. Such a description can be obtained by non-invasive vascular imaging techniques. The evolutionary success of the Computed Tomography Angiography (CTA), in terms of resolution quality, has benefited the clinicians with enhanced image details but at a cost of huge amount of data. Processing of such amount of data is a monotonous, error-prone and time consuming task, which certainly affects the efficiency of the clinicians. Hence, automation of the vessels segmentation in CTA is highly desirable to facilitate a quick and accurate diagnosis.

The tubular shape of the blood vessels offers a great possibility to develop a simple and fast method for vessel segmentation. For instance, the vascular skeleton can first be extracted and used as an initialization step for the vessel segmentation in subsequent stages. A method for fast vascular skeleton extraction was presented in our previous work [1]. In that method, a set of knowledge-based filters is applied to the centeral voxels of potential vessels to distinguish between voxels located inside and outside of the vessel. Voxels that passed through the filters are then connected to generate an approximate vascular skeleton. However, the set of knowledge-based filters also introduces a large number of false positive (FP) nodes. The FPs are further eliminated in the final step of the algorithm - the anatomy-based analysis, which removes most of the spurious branches by examining the shape of their connection.

In this paper, we propose an alternative method for vascular skeleton extraction, where we replace the knowledge-based filters with an efficient convolutional neural networks (CNN) classifier. We evaluate the performance of the CNN classifier using the CTA of the lower limbs in order to compare it with the results obtained by the knowledge-based filters from our previous work. A visual, qualitative, comparison of the resulting skeletons obtained by the two versions of the algorithm is also included.

2 Related work

Due to the large impact of vascular diseases on public health, many scientists are dedicated to research regarding vascular segmentation or vascular centerline extraction. A detailed overview of other vascular segmentation techniques was presented in papers [2, 3]. Here, we briefly review the work that is most relevant for our approach.

Charbonnier et al. [4] recently proposed a method which uses a CNN classifier to improve an airway tree segmentation. In this approach, an initial airway segmentation was provided to classify short airway branch segments into airway or leakage. Each airway candidate was represented by a set of three 2D cross-sectional patches, i.e, the beginning, middle and end of the segment. This set of patches was used as an input for a CNN classifier. Utilizing the CNN classifier significantly improved the quality of a given leaky airway segmentation.

Another method, proposed by Merkow et al. [5] utilized a 3D-CNN to predict the location of the boundary in volumetric data. They demonstrated the performance of their method for the detection of the vascular boundary, but, their approach is not limited to this application. CNN for vessel detection in volumetric images was recently utilized by Gülsün et al. [6]. In their work the blood vessel centerlines were first automatically extracted and then a 1D CNN classifier was used for removing extraneous paths from the detected centerlines. Our proposed method, in comparison, uses the 2D CNN classifier for cross-sectional classification.

3 Dataset

In this work, we utilized 25 CTA volumes of the lower limbs, taken from the clinical routine, to train and validate our proposed method. Initially, four volumes were given to an experienced radiologist for ground-truth labeling. The remaining 21 CTA volumes were kept for independent validation of the method. The radiologist utilized a semi-automatic segmentation tool based on the active contour method, provided by ITK-SNAP [7], to perform the vascular segmentation. The four labeled volumes were used for the detection of the initial nodes and in total provided a set of 352,523 nodes of multi-size vessels and non-vessels.

4 Proposed method

The proposed method is based on the observation that the vessels, being tubular structures, often appear on orthogonal CT slices as bright elliptical-like regions. The method detects voxels located in the middle of such 2D regions, referred to as vessel nodes, and extracts a 2D patch around each node. A CNN classifier is accommodated to classify the patches into two categories, i.e., vessel or non-vessel nodes. Vessel nodes are connected with straight lines, referred to as edges, which results in tree-graph structures. Given a 3D CTA scan of the lower limbs, the method returns one or more tree-graphs representing an approximate skeleton of the vasculature of the lower limbs.

The proposed method for vascular skeleton extraction is a modified version of our previous algorithm [1]. It comprises of four steps that are shown in Fig. 1. This work focuses on improving the nodes classification step. To do so, we replace the previously reported method based on a set of knowledge-based filters with a CNN classifier. Other steps remain the same.

4.1 Node candidates detection

A prerequisite for detecting the bright elliptical regions with the vessel node candidates is the knowledge of the intensity range that corresponds to blood

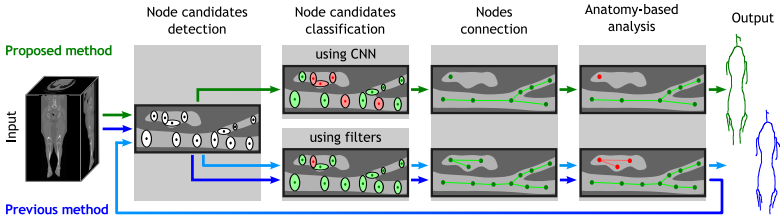


Fig. 1: The pipeline of the proposed method (green, top) produces the final vascular skeleton in one algorithm pass compared to the pipeline of the previous method [1] (blue, bottom) which detect skeletons of larger arteries in the first iteration and adds the skeletons of smaller arteries in the second iteration.

in the input CTA volume. Blood becomes visible on the CTA scan due to the injected contrast medium, which increases its intensity above the intensity of the surrounding muscle tissue. Due to variations in hemodynamics and timing between injection and image acquisition, the blood intensity can differ between different patients and different scans. In [1], we proposed an algorithm based on fitting a sum of Gaussian curves to the image histogram. For each patient input volume, the fitted Gaussian curves automatically define the intensity ranges for three types of tissues: fat, muscle and blood. The intensity range $[\theta^b_{low}, \theta^b_{high}]$ of the blood in each volume is needed for the node candidates detection step and is defined by using this algorithm.

To detect the node candidates, we scan the input volume through all axis oriented planes (axial, sagittal, coronal). Any of the scanned voxels that has an intensity within the range of blood vessel intensities $[\theta^b_{low}, \theta^b_{high}]$ and its position is central within the area of similar intensities, is considered as a node candidate. The central position of the voxel is verified by casting four rays into four main directions starting from the voxel position outwards and confirming that the pair of opposite-pointing rays traversed the same distance until they reached three consecutive voxels with intensities outside of the $[\theta^b_{low}, \theta^b_{high}]$ range. Casting four rays is a sufficient and fast way to verify the central position of the potential nodes. However, not all detected areas are true vessel cross-sections. Due to the partial volume effect, a bone surface, noise, metallic implants or other imaging artifacts, may have intensities similar to blood. Therefore, the detected nodes need to be further classified as either vessel or non-vessel nodes.

4.2 Node candidates classification using CNN

Patch extraction For each node candidate detected in the orthogonal slice, we calculated the biggest diameter and added 2 extra pixels around it to ensure the inclusion of the boundary information. We chose a patch size of 31×31 pixels as an input for the CNN classifier. The patch size of 31×31 pixels was chosen to cover sufficient contextual information of the candidates. At the same time, it provides a good trade-off between a detailed view of the smaller vessels and the possibility to include intermediate and large vessels. The patch pixel values were kept in Hounsfield units, in order not to lose the fine-grained details of the candidates. Some examples of extracted patches are shown in Fig. 2.

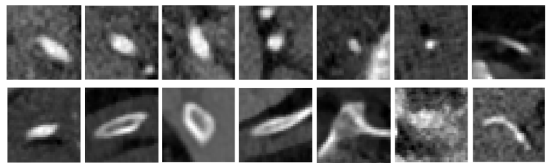


Fig. 2: Some examples of extracted patches of multi-size candidates. The first and the second row shows the patches of vessels and non-vessels, respectively, after intensity normalization.

From all the initially detected candidates, we considered only those having a diameter between 6-27 pixels (ca. 4-20 mm) as a reference set for training the classifier. This resulted in a set of 138,302 potential candidates (24,625 vascular nodes and 113,677 non-vascular nodes). It is not meaningful to train a CNN model on candidates with a diameter smaller than 6 pixels due to insufficient spatial information, leading to inadequate training. Additionally, it is difficult to ensure that the ground truth segmentation of tiny vessels is absolutely correct as it is very tedious if at all possible, to detect all vessels of such small size in lower limb CTA. Candidates having a diameter bigger than 27 pixels were excluded from the reference training set since they were bigger than the chosen patch size (31×31 pixels; including the margin of 2 pixels around the candidate) and hence needed to be resampled to this size. However, both, smaller and larger candidates than the reference candidates, were utilized for testing purposes.

Data partitioning and augmentation We randomly split the reference set into two subsets. One subset was utilized for CNN model development (refer as the model-development subset) and the second one was used for its independent evaluation (refer as the model-evaluation subset). The model-development subset consists of 20,000 samples of each class, whereas the model-evaluation subset was kept imbalanced and contained 4,625 vessels and 93,677 non-vessels samples. The reason for such an imbalanced setting is to evaluate the trained model as per the real clinical scenario, where the frequency of false positives (FP) samples is much higher than the true positives (TP) samples. The model-development subset was again randomly split into training, validation, and testing subsets. Training and validation sets are utilized for cross-validation scheme whereas the test set is utilized for the final model selection. The training set contained 12,000 samples of each class (true vessel nodes and false vessel nodes). The validation and the test sets, both consisted of 4,000 samples of each class.
Generally, the vessels have large variability in terms of contextual surrounding, shape, size, and orientation. Lets assume that such variability can be modeled to a CNN by data-driven approaches. In such way, the classifier (CNN) can learn the orientation-invariant features. However, the number of vessel candidates are usually fewer than the number of non-vessel candidates, which can negatively affect the training of the classifier.We applied several transformations to generate a moderate number of new yet correlated training candidates. Each class of candidates was augmented using the image transformations: horizontal and vertical flipping, translation on x and y axes, and six random angular rotations (0-180°). The translation was limited in moving the candidate position 1 pixel from the center, in order to keep the candidate properly in the patch. This augmentation scheme resulted in 10 augmented variations for each candidate. In such way, classifier will learn the orientation-invariant features. This could be important because the candidates identified by the initial stage are not always centered at the local anatomical structures. Each subset (training and validation) is augmented separately to ensure their independency from each other.

False positive Reduction: CNN configuration The false positive reduction stage is constructed by utilizing a CNN classifier. The architectural design of

our CNN classifier is empirically determined by modifying the network itself. We modified several parameters (i.e. number of layers, kernel size, and types of pooling layer) in a structured way to obtain a better validation accuracy. Amongst all, we further analyzed the usability of two parameters, namely (1) pooling layer and (2) batch normalization (BN). To do so, we developed two models: (1) CNN model with pooling layers and (2) fully-connected convolutional network (FC-CNN) model without pooling layers.
First, we developed a CNN classifier consisting of four convolutional layers, with a max-pooling layer after every second convolution layer. The first convolutional layer consists of 32 kernels of size 3×3 and padded with a two pixels thick frame of zeros. This is done to keep the spatial sizes of the patches same after the first convolutional layer. The second, third and the last convolutional layers consist of 32, 64 and 64 kernels of size 3×3, respectively. The max-pooling layer reduces the size of feature maps by selecting the maximum feature response in overlapping or non-overlapping windows of size 2×2 (stride of 2). The illustration of the CNN model is shown in Fig. 3.

In the FC-CNN model, the pooling layers (maxpooling) were completely removed from the network which resulted in a network of only four convolutional layers. As reported in [12], the FC-CNN could result in an improved performance if the pooling layers are replaced with convolutional layers. In such setting, the network does not lose the spatial representation of the patch. However, such a network can be computationally expensive due to an increased number of network parameters. On the other hand, the feature-wise ordering of the pooling layers can lead to fast optimization, as well as further improve the translation invariance produced by the convolutional layers [12].

In both architectures, we also implemented the recently published Batch Normalization (BN) method [13], after the non-linear (activation function) layers of the network. It normalizes the activations of a feature map for each mini-batch at every optimization step and improves the overall network performance. For the activation function, we utilized the rectified linear units (ReLU) [14] after every convolutional and dense layer. In both networks, the last layer is followed by a dense layer consisting of 512 neurons, which is further connected to the Softmax layer for the final classification into vessels and non-vessels. A comparative evaluation of both models is reported in the result section.

Training Before feeding the training data to the network, we normalize the intensity for each patch by subtracting the mean and dividing by the standard deviation. In such a way, the uneven distribution of intensities is scaled into a normalized intensity distribution and lead to a better convergence. The same procedure was applied during testing. The classifier was trained in a 5-fold cross-validation scheme. The candidates were randomly split into five blocks to ensure that each set was utilized as validation set once in each fold.

The RMSProp [8] is used to efficiently optimize the weights of the CNN. It normalizes the gradients by utilizing the magnitude of recent gradients. The weights are initialized as proposed in [9] and updated in a mini-batch scheme of 128 candidates at a rate of 0.001. The biases were initialized with zero. A

dropout [10] of 0.25 is implemented on the output of each pooling layer and a dropout of 0.5 is implemented on the output of the dense layer. Softmax loss is utilized to predict the final output. The CNN model is implemented using Theano backend in Keras [11]. The training continued for 20 epochs with an average training time of 103 seconds/epoch on a GPU GeForce GTX 680.

4.3 Nodes connection

The nodes that were classified by the CNN classifier as true vessel nodes are, in this step, linked together by using simple connection rules. First, each node is considered a separate graph. A link between two nodes is established, if the two nodes are close neighbors, all voxels on the line connecting these two nodes have an intensity within the range $[\theta^b_{low}, \theta^b_{high}]$ and these two nodes were not yet connected via other nodes. After all possible connections between the nodes have been created the preliminary vascular tree-graph structure is obtained which needs to be cleaned from possible spurious branches.

4.4 Anatomy-based analysis

The anatomy-based analysis step cleans the preliminary tree-graph structure from spurious non-vessel graphs or graph segments. The cleaning is based on the observation that vessel nodes are connected into straight or slightly corrugated branches whereas non-vessel nodes, often arising on the bone surface, are linked into unorganized and zigzag branches. Calculating the average angle between the line segments per branch allows distinguishing between true and false graph segments depending on the average angle being greater than 135° or not, respectively. The anatomy-based analysis step also closes small gaps between two segments and removes very short graphs (for details see [1]). Finally, it returns clean tree-graph structure that corresponds to the approximate vascular skeleton.

5 Evaluation and Results

We compare the performance of two architectures, CNN and FC-CNN. Subsequently, the performance of the whole algorithm for vascular skeleton extraction,

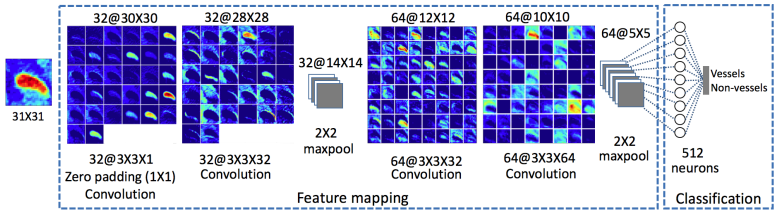


Fig. 3: An overview of the proposed CNN classifier, showing the output of each convolution filter applied to an example patch of a vessel. Here, the greyscale intensities are shown in color for suitable visualization.

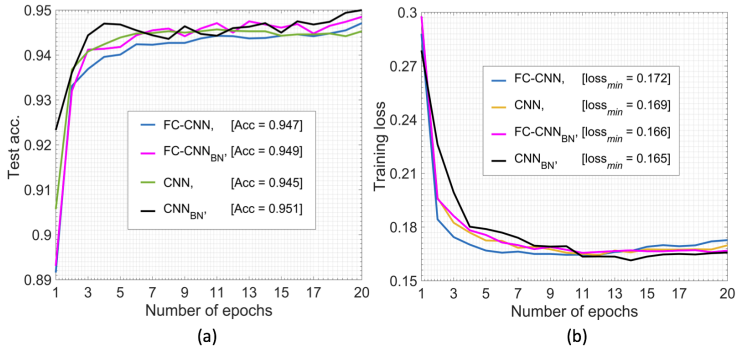


Fig. 4: Performance curves of different configurations at different number of epochs: (a) Test accuracy, (b) Training loss of both networks with and without batch normalization (BN) method.

using the proposed method is qualitatively evaluated and compared to the previous skeleton extraction method [1].

5.1 CNN versus FC-CNN

A comparative performance of both networks with and without BN is shown in the Fig. 4(a)-(b). The left figure shows the test accuracy of each configuration at different number of epochs on the test subset of the model development set (consisting 4000 samples of each class). It is noticeable that both networks with BN yield a better accuracy in comparison to the networks without BN. Interestingly, in the case of non-batch normalization configuration, the FC-CNN classifier also achieves a higher accuracy than the CNN classifier. Both networks trained with BN resulted in a better training loss. These results are in line with similar findings on the original BN work [13]. Comparatively, the CNN, trained with BN, resulted in better accuracy as the FC-CNN classifier (validated by paired t-test) with less number of parameters. The FC-CNN and CNN classifiers consist of total ca. 30.1 million and 1.2 million parameters, respectively. Therefore, we decided to utilized the CNN classifier for our application.

5.2 CNN and knowledge-based filters evaluation

Knowledge-based filters The knowledge-based filters, proposed in [1] are simple filters derived from the characteristic appearance of vessels. They quickly remove the node candidates that do not fulfill the vessel characteristics. These filters examine: (1) if the artery lumen is homogeneously filled with blood, (2) if the inside of the artery is brighter than the outside, (3) if the vascular cross-section has regular elliptical or circular shape and (4) if the close artery neighborhood contain only intensities corresponding to fat, muscles or blood. The reason for the last filter is that a vessel can be adjacent to either fat, muscles or bones, but

the partial volume effect will cause the surface of the bone to have decreased intensities that overlap with blood intensities.

Quantitative evaluation Quantitative evaluation was performed on the model-evaluation subset (4,625 vessels and 93,677 non-vessels) of patches not used in the model development process. The performance of the proposed CNN model and knowledge-based filters was evaluated on this subset in terms of *Precision*, *Recall*, and *F-score*. Additionally, for the CNN model, the *Area under the Precision-Recall curve (AUC)* is also presented in Fig. 5. The evaluation was performed separately for small vessels (< 4 mm), medium-sized vessels (4-20 mm) and large vessels (> 20 mm). This division is motivated by the fact that the CNN classifier was trained for the middle-sized vessels, however, it was used to classify small and large vessel candidates as well. Table 1 shows the resulting values for each evaluation measure per classifier and per candidate group.

Qualitative evaluation Qualitative evaluation was performed by visual comparison of the resulting skeletons extracted from 21 CTA volumes of lower limbs by using two pipelines, the proposed algorithm pipeline and the pipeline presented in our previous work [1]. The schematic illustration of the pipelines in Fig. 1, shows that the previous filter-based algorithm needs to run in two iterations. The skeleton of large vessels, extracted in the first iteration serves as a basis for distinguishing between true and spurious graphs of small arteries detected in the second iteration. Small artery candidates are easier to mistake for noise or candidates detected on the bone surface and a useful indication about their true belongingness is the connection to the graph established in the first iteration. Since the proposed CNN classifier improves the false positive candidate

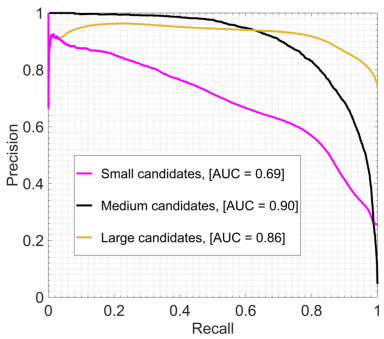
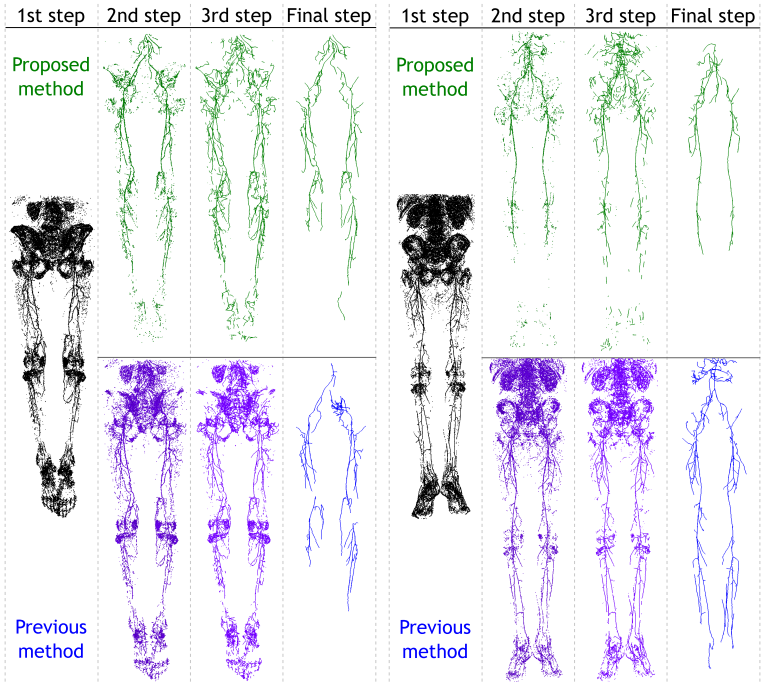


Fig. 5: Precision-Recall curves showing the performance of CNN classifier for the test subset of different sizes of candidates.

	Set	Prec.	Rec.	F-score	AUC
CNN	small	0.66	0.71	0.65	0.69
	medium	0.81	0.83	0.82	0.90
	large	0.70	0.75	0.72	0.86
filters	small	0.29	0.78	0.42	–
	medium	0.28	0.89	0.43	–
	large	0.06	0.58	0.11	–

Table 1: Comparative evaluation of CNN classifier and knowledge-based filters.



(a) Volume A (b) Volume B

Fig. 6: Results after each algorithm step for 2 volumes; result after the 1st step is the same for both methods.

rate, having two iterations is not relevant anymore and it is possible to simplify the pipeline while still obtaining better results.

It is crucial to determine a suitable threshold level for the decision boundary of the CNN classifier. We experimented with several threshold values ranging from 0.9 (corresponds to high precision) to 0.45 (corresponds to high recall). We observed that the skeleton, resulting from the higher thresholds, rarely contains the spurious branches and does not need the further anatomy-based analysis step. However, it also reduces the number of true vessel candidates and leads to fewer vessel branches. On the other hand, a lower threshold value detects a larger number of false positives and many spurious branches. This resembles the behavior of the simple knowledge-based filters. After this empirical analysis, we finally decided to select 0.65 as the suitable threshold level for our application.

From 21 resulting skeletons, we selected two representative results that demonstrate the merits and demerits of both methods. Figure 6 shows a comparison of the final skeletons along with the partial results obtained after each algorithm step by both algorithms. The results, after the second step, confirmed that the CNN classifier improves the FP rate and keeps a fewer number of candidates from which the majority are true vessel candidates. The final results of volume A (Fig. 6a) show that the proposed method detects more vascular branches, and compared to the previous method does not contain spurious graph segments in the pelvis region. On the other hand, we also observed some cases where the CNN classifier discarded a larger amount of true vessel candidates, which led to missing a complete vessel branch. The final results of volume B depict an example of such a case. After an in-depth analysis of these vessels, we noticed that these vessels were either very small or they were very diseased. The CNN classifier was not trained for classifying such small vessels, which explains the lower performance in case of small vessels. In a case of diseased vessels, there exist many different variations between the appearance of diseased vessels depending on the type and the seriousness of the disease. Our dataset is taken from a clinical routine and contains large variations between the patient material. Therefore, using patches of four volumes in the training process was not sufficient enough to cover all the possible clinical variability. In the case of volume B, the knowledge-based filters were more inclusive in keeping true and false candidates, which resulted in a better skeleton compared to the proposed method.

Computation time The computation time needed to process a set of 130,000 patches which is approximately the average number of patches per volume in our dataset, was measured for both classifiers. The knowledge-based filters took ca. 22 seconds to process the patches, whereas the CNN classifier took ca. 30 seconds.

6 Discussion and Conclusions

Our main goal to reduce the false positive rate by using a CNN classifier was successfully fulfilled. Both the quantitative and qualitative evaluations support this. In comparison to our previous work, the detection of vessel nodes can now be processed in a single iteration, resulting in a simplified methodological pipeline.

Interestingly, in some cases, our previously proposed simple knowledge-based filters in combination with the anatomy-based analysis step, also performed well in comparison to the proposed method. One reason for the occasional lower performance of the proposed method could be the high variance in the diseased arterial vessels of each patient. To improve on this, the network could be remodeled with a bigger training set of vessels with a wider variation of arterial diseases. To do so, a reliable and consistent labeling of ground truth segmentation is desired, which is a tedious and difficult task; especially for the small and tiny vessels.

In order to further improve the performance, multiple patches from the different axes oriented planes can be utilized to remodeled the CNN model. Alter-

natively, the 3D clusters of the candidates can be extracted to train a 3D CNN model for classification.

Acknowledgements

Lidayová, Frimmel, Bengtsson, and Smedby have been supported by the Swedish Research Council (VR), grant no. 621-2014-6153. Gupta has been supported by Skype IT Academy Stipend Program, EU institutional grant IUT19-11 of Estonian Research Council.

References

1. Lidayová, K., Frimmel, H., Wang, C., Bengtsson, E., Smedby, Ö.: Fast vascular skeleton extraction algorithm. Pattern Recognition Letters, 76, 67–75 (2016)
2. Kirbas, C., Quek, F.: A review of vessel extraction techniques and algorithms. ACM Computing Surveys (CSUR), 36(2), 81–121, (2004)
3. Lesage, D., Angelini, E. D., Bloch, I., Funka-Lea, G.: A review of 3D vessel lumen segmentation techniques: Models, features and extraction schemes. Med. Image Anal., 13(6), 819–845, (2009)
4. Charbonnier, J. P., van Rikxoort, E. M., Setio, A. A., Schaefer-Prokop, C. M., van Ginneken, B., Ciompi, F.: Improving airway segmentation in computed tomography using leak detection with convolutional networks. Med. Image Anal., 36, 52–60. (2017)
5. Merkow, J., Marsden, A., Kriegman, D., Tu, Z.: Dense Volume-to-Volume Vascular Boundary Detection. In International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer International Publishing, 371–379. (2016)
6. Gülsün, M. A., Funka-Lea, G., Sharma, P., Rapaka, S., Zheng, Y.: Coronary Centerline Extraction via Optimal Flow Paths and CNN Path Pruning. In International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer International Publishing, 317–325. (2016)
7. Yushkevich, P. A., Piven, J., Hazlett, H. C., Smith, R. G., Ho, S., Gee, J. C., Gerig, G.: User-guided 3D active contour segmentation of anatomical structures: significantly improved efficiency and reliability. Neuroimage, 31(3), 1116–1128. (2006)
8. Tieleman, T., Hinton, G.: Lecture 6.5-RmsProp: Divide the gradient by a running average of its recent magnitude. COURSERA: Neural Networks for ML. (2012)
9. Glorot, X., Bengio, Y.: Understanding the difficulty of training deep feed-forward neural networks. In: Aistats. vol. 9, 249–256. (2010)
10. Srivastava, N., Hinton, G.E., Krizhevsky, A., Sutskever, I., Salakhutdinov, R.: Dropout: a simple way to prevent neural networks from overfitting. Journal of Machine Learning Research 15(1), 1929–1958. (2014)
11. Chollet, F.: Keras. https://github.com/fchollet/keras. (2015)
12. Springenberg, J.T., Dosovitskiy, A., Brox, T. and Riedmiller, M.: Striving for simplicity: The all convolutional net. Proceedings of the 3rd International Conference on Learning Representations (ICLR). (2015)
13. Ioffe, S. and Szegedy, C.: Batch normalization: Accelerating deep network training by reducing internal covariate shift. arXiv preprint arXiv:1502.03167. (2015)
14. Nair, V., Hinton, G.E.: Rectified linear units improve restricted boltzmann machines. In: 27th International Conference on Machine Learning, 807–814. (2010)

CURRICULUM VITAE

Personal data

Name:	Anindya Gupta
Date of birth:	December 21, 1988
Place of birth:	Jaipur, India
Citizenship:	India

Language Competence

Hindi:	Native
English:	Proficient

Contact data

Address:	Akadeemia tee 5, 115, Tallinn, Estonia
Phone:	+372 56984664
E-mail:	anindya.gupta@ttu.ee

Education

2014 – 2018: (expected)	Tallinn University of Technology, Tallinn, Estonia	PhD, Electronics and Telecommunication
2011 – 2013:	Kingston University, London, England	MSc in Software Engineering
2006 – 2010:	Rajasthan Technical University, Rajasthan, India	B.Tech. in Computer Science
2004 – 2006:	Kendriya Vidyalaya, Rajasthan, India	Certificate of Senior Secondary

Professional Experience

2016 – 2017:	Center for Image Analysis, Uppsala University, Sweden	Visiting Ph.D. Scholar (Multiple visits)
2012 – 2013:	Digital Info. Research Centre, Kingston University, England	Research Assistant
2010 – 2011:	Genpact Pvt. Ltd., Jaipur, India	Data Analyst

Invited Talks

"Multi-layer Perceptron for Pulmonary Nodules Detection in CT Images and Cilia Detection in Low Magnification TEM Images," at the Center for Image Analysis, Uppsala University, Sweden on November 11, 2016.

"Deep neural networks for the classification and denoising of medical and biomedical images," at the Center for Image Analysis, Uppsala University, Sweden on January 29, 2018.

Fellowships and Awards

2014 – 2018:	Skype IT Academy Grants for Excellence, Estonia
2014 – 2015:	Research Stipend from Eliko Competence Centre, Estonia
2016 – 2017:	Erasmus Mundus Mobility Grant for Exchange Visit
2017 – 2018:	Dora Plus Ph.D. Mobility Grant for Exchange Visit

ELULOOKIRJELDUS

Isikuandmed

Nimi:	Anindya Gupta
Sünniaeg:	Detsember 21, 1988
Sünnikoht:	Jaipur, India
Kodakondsus:	India

Keeleoskus

Hindi:	emakeel
Inglise keel:	suurepärane

Kontaktandmed

Aadress:	Akadeemia tee 5, 115, Tallinn, Eesti
Telefon:	+372 56984664
E-post:	anindya.gupta@ttu.ee

Hariduskäik

2014 – 2018: (expected)	Tallinna Tehnikaülikool, Tallinn, Eesti	Doktorantuur, elektroonika ja telekommunikatsioon
2011 – 2013:	Kingstoni ülikool, London, Ühendkuningriik	Tarkvaratehnika magister
2006 – 2010:	Rajasthani tehnikaülikool, Rajasthan, India	Arvutiteaduse bakalaurus
2004 – 2006:	Kendriya Vidyalaya, Rajasthan, India	Keskkooli lõputunnistus

Teenistuskäik

2016 – 2017:	Kujutiseanalüüsikeskus, Uppsala Ülikooli, Rootsi	külalisdokoktorand (mitmed külastused)
2012 – 2013:	Digitaalse Info uurimiskeskus, Kingstoni ülikool,	uurimisassistent
2010 – 2011:	Genpact Pvt. Ltd., Jaipur, India	Andmeanalüütik

Kutsutud loengud

"Multi-layer Perceptron for Pulmonary Nodules Detection in CT Images and Cilia Detection in Low Magnification TEM Images," at the Center for Image Analysis, Uppsala University, Sweden on November 11, 2016.

"Deep neural networks for the classification and denoising of medical and biomedical images," at the Center for Image Analysis, Uppsala University, Sweden on January 29, 2018.

Tunnustused ja stipendiumid

2014 – 2018:	Skype IT Akadeemia Stipendiumid Väljapaistvate Tulemuste Eest
2014 – 2015:	Eliko kompetentsikeskuse teadustööstipendium, Eesti
2016 – 2017:	Erasmus Mundus Mobiilsusgrant
2017 – 2018:	Dora Plus Mobiilsusgrant