Spring 2018

Detecting cells and analyzing their behaviors in microscopy images using deep neural networks

Yunxiang Mao

Department: Computer Science

Recommended Citation

Mao, Yunxiang, "Detecting cells and analyzing their behaviors in microscopy images using deep neural networks" (2018). *Doctoral Dissertations*. 3135.
https://scholarsmine.mst.edu/doctoral_dissertations/3135

DETECTING CELLS AND ANALYZING THEIR BEHAVIORS IN MICROSCOPY IMAGES USING DEEP NEURAL NETWORKS

by

YUNXIANG MAO

A DISSERTATION

Presented to the Graduate Faculty of the

MISSOURI UNIVERSITY OF SCIENCE AND TECHNOLOGY

In Partial Fulfillment of the Requirements for the Degree

DOCTOR OF PHILOSOPHY

in

COMPUTER SCIENCE

2018

Approved by

Zhaozheng Yin, Advisor
Wei Jiang
Dan Lin
Yanjie Fu
Ruwen Qin

ABSTRACT

The computer-aided analysis in the medical imaging field has attracted a lot of attention for the past decade. The goal of computer-vision based medical image analysis is to provide automated tools to relieve the burden of human experts such as radiologists and physicians. More specifically, these computer-aided methods are to help identify, classify and quantify patterns in medical images. Recent advances in machine learning, more specifically, in the way of deep learning, have made a big leap to boost the performance of various medical applications. The fundamental core of these advances is exploiting hierarchical feature representations by various deep learning models, instead of handcrafted features based on domain-specific knowledge.

In the work presented in this dissertation, we are particularly interested in exploring the power of deep neural network in the Circulating Tumor Cells detection and mitosis event detection. We will introduce the Convolutional Neural Networks and the designed training methodology for Circulating Tumor Cells detection, a Hierarchical Convolutional Neural Networks model and a Two-Stream Bidirectional Long Short-Term Memory model for mitosis event detection and its stage localization in phase-contrast microscopy images.

ACKNOWLEDGMENTS

I would like to express my gratitude to all those who have helped me during this research. Firstly, I would like to thank my advisor Dr. Zhaozheng Yin for giving me the opportunity to do this research, his valuable insights and suggestions have helped me to overcome many hurdles during this work. I am grateful to him for the advice and guidance given by him throughout my PhD program. I would also like to thank Dr. Wei Jiang, Dr. Dan Lin, Dr. Yanjie Fu and Dr. Ruwen Qin for accepting to be on my PhD committee and for their time to review this work.

Further, I would like to extend my thanks to the Department of Computer Science and the Intelligent System Center (ISC) at Missouri S & T, and the National Science Foundation (NSF) for supporting my PhD program. I would like to thank Dr. Ming Leu and the staff at Engineering Research Laboratory (ERL) at Missouri University of Science and Technology for their assistance in providing me a nice place to do my research.

My sincere thankfulness also goes to my colleagues in Dr. Zhaozheng Yin's research team. They are my best friends in my PhD life, my comrades in arms who accompanied me in the last 5 years of study.

Last but not the least, a deep sense of gratitude goes to my family for their endless support for my endeavors on the road to pursue my dreams.

TABLE OF CONTENTS

Page

SECTION

LIST OF ILLUSTRATIONS

Figure Page

LIST OF TABLES

Table Page

1. INTRODUCTION

Over the past few decades, medical imaging techniques, such as computed tomography (CT), magnetic resonance imaging (MRI), positron emission tomography (PET), mammography, ultrasound, and X-ray, have been used for the early detection, diagnosis, and treatment of diseases. In the clinic, medical images are interpreted mostly by human experts such as radiologists and physicians. However, given wide variations in pathology and the potential fatigue of human experts, researchers and doctors have begun to benefit from computer-aided interventions.

In the early stage of medical image analysis, from the 1970s to 1990s, low-level pixel processing (e.g., edge and line detector filters, region growing) and mathematical modeling (e.g., fitting lines, circles and ellipses) are used to construct rule-based systems to solve particular tasks. Later on at the end of the 1990s, supervised methods, where training data are used to build a system, became increasingly popular in medical image analysis. The feature extraction and statistical classifiers form the concept of pattern recognition and machine learning. Thus the systems that are completely designed by humans shifted to systems that are trained by computers using training data from which the feature representations are extracted. A key point of the design of such systems is finding or learning informative features that well describe the patterns in images or videos. Those handcrafted features are only able to find informative patterns inherent in data at a shallow level, thus limiting their representational power.

The optimal features representations should not be designed by human, but automatically learned by the models. Deep learning has overcome the drawbacks of handcrafted features by discovering the high-level representative features in a self-taught manner. One of the most successful types of models for image analysis so far is convolutional neural networks (CNNs). A typical CNNs model consists of several layers with convolution filters which extract features from low level to high level. The first successful real-world application is LeNet by [1] for hand-written digit recognition. One ground-breaking work is the contribution of Krizhevsky et al. (2012) to the ImageNet challenge. In subsequent years after that, CNNs have become the an effective model for different tasks in computer vision.

The medical image analysis community has noticed these developments. Systems designed to solve particular problems in the medical field have transited from handcrafted features to deep learned features. Bengio et al. [2] provide a thorough review of previously handcrafted-feature based techniques. They include principal component analysis, clustering of image patches, dictionary approaches, and many more. Shen et al. [3] provide a review on application of deep learning to medical image analysis. Litjens et al. [4] offer a comprehensive overview of almost all fields in medical imaging. Among all the tasks in medical image analysis, this dissertation focuses on developing deep learning methods to detect specimens in microscopy images and understand their behavior with two case studies:: (1) Circulating Tumor Cell detection, and (2) mitosis cell detection and its stage localization, with the power of deep learning.

1.1. CIRCULATING TUMOR CELL DETECTION

The number of Circulating Tumor Cells (CTCs) in blood provides an indication of disease progression and tumor response to chemotherapeutic agents. Hence, routine detection and enumeration of CTCs in clinical blood samples have significant applications in early cancer diagnosis and treatment monitoring. During the process of cancer metastasis, malignant cells may break away from the primary tumor, enter the blood or lymphatics system, and then form a secondary tumor at a distant organ site. A number of studies have shown that in the early stages of solid tumor progression, before overt clinical signs, malignant cells are found circulating in blood [5, 6, 7, 8]. The number of Circulating Tumor Cells (CTCs) in blood is a predictor of disease progression and an indicator of tumor response to chemotherapeutic agents [7, 8]. Thus, there is much interest in the development of routine techniques for detection and enumeration of CTCs in clinical blood samples. An automated technique for enumeration of CTCs may aid in the early diagnosis of cancer even before tumors are visible using traditional imaging approaches and may help avoid the use of more invasive techniques such as tumor biopsy.

1.2. MITOSIS DETECTION

Analyzing the proliferative behavior of stem cells in vitro plays an important role in many biomedical applications, such as stem cell manufacturing, drug discovery and tissue engineering. Accurate enumeration and localization of the occurrences of *mitosis*, which is the process whereby the genetic material of a eukaryotic cell is equally divided, resulting in daughter cells, are critical to monitor the health and growth rate of cells. For small scale research studies, manually enumeration and localization of mitosis event by histologists may be considered. However, when it comes to large-scale research, analysis of these images becomes an arduous process which involve many hours of human inspection. Traditional methods for measuring cell proliferation have been developed for many years. Most of the analysis methods use fluorescent, luminescent or colorimetric microscopy images which are acquired by invasive methods, such as staining cells with fluorescent dyes and radiating them with the specific wave-length light. The invasive method damages cells' viability or kills cells, which does not allow continuously monitoring the cell proliferation process. Phase-contrast microscopy, as a non-invasive imaging modality, offers the possibility to persistently monitor cells' behavior in the culturing dish without altering them [9]. Thus *mitosis detection* in microscopy images is the direction of cell behavior analysis.

In fact, the computer vision based mitosis detection contains two tasks: (1) mitosis event localization and (2) mitosis stage localization. Given a microscopy image sequence, the mitosis event localization refers to identify where and when mitosis happen in the sequence. For the mitosis stage localization, we are trying to identify different mitotic phases within the mitotic sequences. Accurately localizing the time of each stage will facilitate the quantification of biological metrics, allowing biologists to assess different factors that impact the length of time a cell spends in each stage of the mitosis.

1.3. MOTIVATIONS AND CONTRIBUTIONS

Methods that use antibodies against tumor cells require prior knowledge of the markers which vary widely according to different types and stages of cancer. These antibody-based systems that efficiently detect carcinomas will miss many other types of malignancies including leukemia, lymphoma and non-epithelial tumors. Svensson et al. [10] use a Bayesian classifier based on a probabilistic generative mixture model to detect CTCs. However, their system is based on fluorescent microscopy images which is an invasive approach. *There is a great interest in the CTC community to develop an noninvasive method that is not dependent on tumor cell markers and capable of detection across a wide range of cancer types.*

A phase contrast microscopy image containing some CTCs is shown in Fig.1.1(a). The CTCs exhibit large variations in shape and size and they overlap with each other (Fig.1.1(b)). Some non-CTC background has similar appearance to the CTCs (Fig.1.1(c)). It is very hard to distinguish CTCs from the background by simple intensity thresholding or morphological operation.

Therefore, *reliable image features are needed to detect CTCs.* Deep Convolutional Neural Network (DCNN) has shown its effectiveness on object detection and classification in recent years [11, 12]. It has been proved to be an effective tool in several biomedical applications such as mitosis cell detection [13, 14]. In a DCNN architecture, it has several layers of convolutional filters with each layer followed by either max or mean pooling operations to produce abstract and useful representations of the input object. The parameters

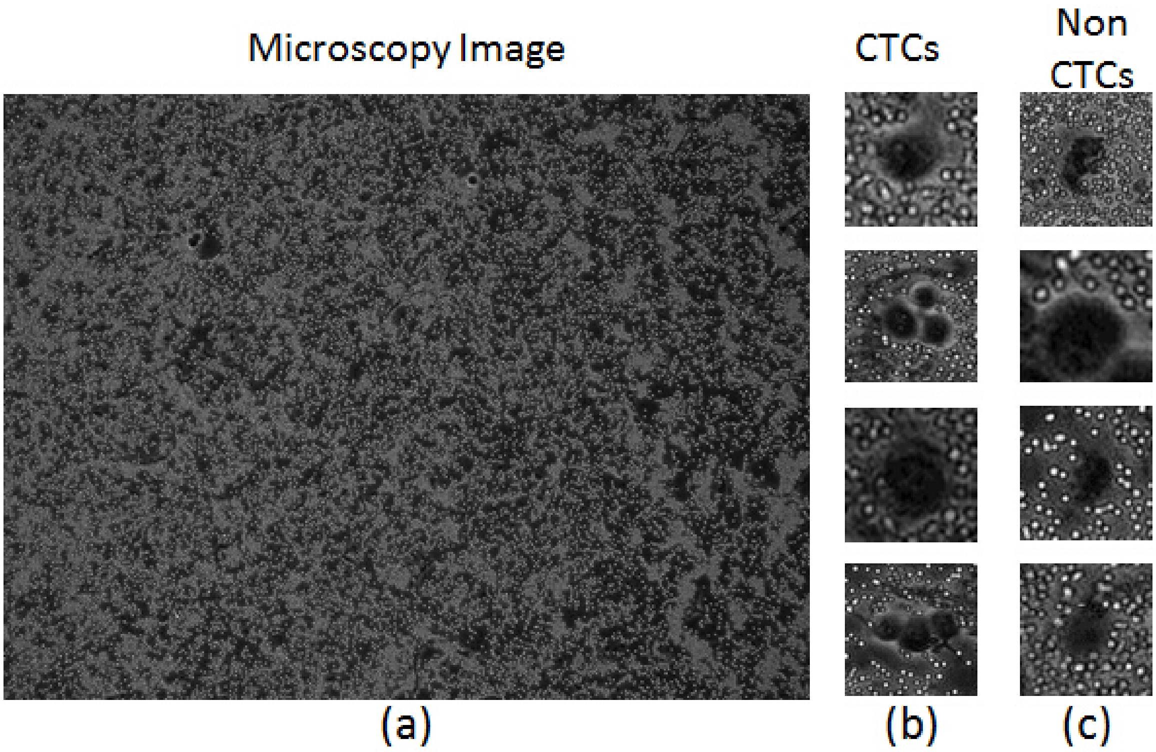


Figure 1.1. Visualize Circulating Tumor Cells using phase contrast microscopy imaging.

of kernals are automatically learned without any human effort. *The learned convolutional kernels are more effective feature extractors compared to other human-designed feature descriptors.*

The balance between the number of positive and negative samples is quite important in the DCNN training. However, CTCs in blood are infrequent so that it is hard to acquire a large amount of CTC image patches for DCNN training. Meanwhile, there are much more negative samples from the background with redundancies so that it is infeasible to include every possible negative sample in training. Thus, a *training methodology which is able to collect the most representative training samples from limited training images is needed to avoid the class imbalance problem.*

The above three needs (noninvasive microscopy imaging, image feature extraction, and training with representative samples) motivate us to develop a CTC detection system.

As for the mitosis detection, most of the previous mitosis detection approaches either use handcrafted features or consider a single image as the input of DCNN architectures. If we attempt to detect mitosis events by a single image, we may lose the visual appearance change context during the whole process of a mitosis event. Furthermore, motion information hidden in the continuous image sequence can also aid the detection of mitosis event. In the work of Su et al. [15], 3D objects recognition is achieved by a multi-view CNN. Each single CNN in the first layer takes an image of the object captured in one aspect as input. After the first-layer CNNs, and a pooling operation and a single CNN is adopted to converge the features to predict the label of object. Inspired by their work, we propose a Hierarchical Convolutional Neural Network (HCNN) for the task of mitosis event detection, which utilizes the temporal appearance change information and motion information in continuous microcopy images.

Furthermore, another task in mitosis detection is to localize each stage of a mitosis process. The HCNN take fixed-size input and output one label for the patch sequence while the length of extracted sequences varies, it is not able to perform the task of localization of different mitosis stages. Long-term Short Memory (LSTM) [16], which is able to address variant-length input, is widely used in natural language processing. It can be adapted to many-to-many, many-to-one, and one-to-many models according to different tasks [17]. Hence, to conquer to drawback of HCNN, we further propose a LSTM-based model which take the entire patch sequences as the input to detect mitosis in the input sequence.

In this dissertation, we are solving the problems of Circulating Tumor Cells (CTCs) detection and mitosis event detection. Our contributions are summarized below.

- We propose an image-based CTC detection system based on DCNN. The proposed system is non-invasive without staining markers that damage the viabilities of CTCs.
- An effective training methodology is proposed. It finds the most representative samples to better define the classification boundary between positive and negative samples.

- We propose a Hierarchical Convolutional Neural Network (HCNN) to classify each candidate patch sequence based on its temporal appearance and motion information.
- We design a Two-stream Bidirectional Long-Short Term Memory (TS-BLSTM) which is able to localize four stages of the mitosis sequence patches.

1.4. ORGANIZATION OF DISSERTATION

The rest of the dissertation is organized as follows. In Section 2, related work is discussed. In Section 3, a Convolutional Neural Network is designed to detect the CTCs. An iteratively training algorithm is proposed to reduce the training time. A round-based training method is proposed to further improve the performance of detection by finding and training on the most representative samples. Section 4 presents a Hierarchical Convolutional Neural Networks model to solve the problem of mitosis detection. To solve the mitosis event detection and its stage localization, Section 5 introduces a Two-Stream Bidirectional Long Short-Term Memory model to solve these two problem simultaneously. Finally Section 6 concludes the work.

2. LITERATURE REVIEW

2.1. TRADITIONAL CTC DETECTION

Because CTCs in blood are infrequent, 1 per 1 billion normal cells found in the blood [5], routine detection of CTCs poses a significant challenge. Several methods have been developed to quantify and capture CTCs from human blood. Many of these approaches depend on surface markers expressed on tumor cells which can be exploited for use in positive selection. One widely used marker for detection of carcinoma cells in the blood is epithelial cell adhesion molecule (epCAM) [18]. In some enrichment techniques magnetic beads with immobilized anti-epCAM [19], and other anti-tumor antibodies [20], are used for immunomagnetic separation of malignant cells from the normal blood cell population. Immunomagnetic-based selection of CTCs is attractive because of its simplicity and the availability of the needed tools and reagents. Commercially available systems, based on epCAM-positive selection, have been successfully implemented in CTC evaluation in certain types of carcinoma. Other approaches exploit differences in tumor cell physical properties such as size, density or adhesiveness [21]. To the best of our knowledge, little work has been done on image-based CTC detection.

2.2. TRADITIONAL MITOSIS DETECTION

Several tracking-based mitosis detection on phase-contrast microscopy images methods have been proposed in the past decade. Debeir et al. [22] combined several model-based mean-shift processes to track migrating cells. In [23], Padfiled et al. segment the nuclei with a shape/size constraint and use an Euclidean distance metric to link different phases of the cell cycle. AL-Kofahi et al. [24] segment cells in each image of the sequence, and adopt a multiple-object matching method which measures a number of cell attributes such as size, shape and location to track cells. Li et al. [25] exploit a geometric active contour model to track detected cells. Liang et al. [26] segmented the cell nuclei from background, tracked the nuclei as cell sequences and then utilized a conditional random field (CRF) model [27] with shape and texture features of the segmented nuclei to identify mitosis event. The problem of mitosis detection in these papers is solved based on volumetric image segmentation or object tracking algorithms with the goal of tracking cell movements over time. The mitoses are identified based on the temporal progression of cell features or the connection between the segmented mother and daughter cells. However, these mitosis event detection approaches heavily depend on the long-range object tracking performance, which itself is a very challenging task.

Considering the drawback of tracking-based mitosis detection, tracking-free approaches detect mitosis directly in an image sequence. Huang et al. [28] propose an algorithm called eXclusive Independent Component Analysis (XICA) which focuse on the components of differences between two classes of training patterns rather than the major components. They classify the given testing pattern by computing the residual of the relative exclusive basis set. Since the mitosis is a dynamic process, the performance of mitosis detection would benefit from taking advantage of the temporal dynamic information in the evolution of visual patterns. These methods usually consist of three steps: candidate detection, feature extraction and classification. In candidate detection, which aim to produce image patches which contain mitosis event, thresholding and/or morphological operations are typically applied. Gallardo et al. [29] adopted a hidden Markov model (HMM) to classify candidates based on temporal patterns of cell shape and appearance features. Li et. al [25] apply a cascade classifier framework [30] to classify volumetric sliding windows of an image sequence based on 3D haar-like features [31]. The proposed method is efficient since it only needs one sequential scan through the image sequence with a trained classifier. However, the coarse resolution of the sliding temporal window may limit the localization precision.

Liu et al. [32] proposed an approach based on Hidden Conditional Random Fields (HCRF) [33] in which mitosis candidate patch sequences are extracted through a 3D seeded region growing method, then HCRF is trained to classify each candidate patch sequence. This method does not rely on object tracking algorithms and achieves good performance on C3H10T1/2 stem cell datasets. Since only one label is assigned to a patch sequence, this HCRF-based approach is able to identify if a patch sequence contains mitosis or nor, but it can not accurately localize the birth moment of the mitosis event in the patch sequence.

A few extensions have been made on the HCRF-based approach. Huh et al. [34] proposed an Event-Detection CRF (EDCRF) in which each patch in a candidate sequence is assigned with one label. The birth moment of the mitosis event is determined based on the observation that if there exists a change from “before mitosis” to “after mitosis” label. Liu et al. [35] utilized a maximum-margin learning framework for training the HCRF and proposed a semi-Markovian model to localize mitosis events.

2.3. DEEP-LEARNING BASED MITOSIS DETECTION

The previous approaches rely on handcrafted image features. Deep Convolutional Neural Network (DCNN) which is capable of learning feature representations from big data and modeling the large variation among the data, has shown its effectiveness on object detection and classification. The learned kernels in DCNN are very effective feature extractors compared with handcrafted feature extractors. In order to improve the performance of DCNN on the challenging ImageNet dataset, Yan et al. [36] propose a hierarchical deep convolutional neural network which classify the input images with several components within the architecture. The low-level features of input image are extracted through the lower layers, and then the classification of input image is done followed by a coarse-to-fine approach. Wang et al. [37] proposed a hybrid deep learning network for the task of face verification. The input to their deep model is multiple pairs of different subregions in the two original images to be compared. The final output will be binary which indicates whether the two original images come from the same person. Cireşan et al. [13] utilized the DCNN as a pixel classifier for mitosis detection in individual breast cancer histology images. During the histology, the histologic specimens are stained and sandwiched between a glass microscope slide and coverslip. So this DCNN method is not suitable for detecting a continuous mitosis event in the time-lapse phase-contrast microscopy image sequences. The latest CNN-based methods [38, 39] achieve good performance on the task of mitosis event detection. Mao et al. build a hierarchical convolutional neural networks in which both appearance and motion temporal information are utilized to detect the birth moment of a mitosis sequence. Compared with traditional CNN which only take one single image as input, 3D CNN [40, 41] extract temporal features though its 3D convolutional kernels. Wei et al. [39] design several different 3D CNN architecture and demonstrate that 3D CNN outperform the 2D CNN features and other hand-crafted features.

The previous CNN-based methods only accept a fixed-size vector as input and produce a fixed-size vector as output, e.g. probabilities of different classes. However, the length of extracted sequences varies. Furthermore, since their models take fixed-size input and output one label for the patch sequence, they are not able to perform the task of localization of different mitosis stages. Long-term Short Memory (LSTM) [16], which is able to address variant-length input, is widely used in natural language processing. It can be adapted to many-to-many, many-to-one, and one-to-many models according to different tasks [17]. For the task of mitosis stage localization, the many-to-many model can be utilized to output one label for each image in the candidate sequence to label its stage.

3. CIRCULATING TUMOR CELLS DETECTION

3.1. DATA ACQUISITION

We labeled MCF-7 breast cancer cells with a red fluorescence cell-tracker dye for 30 minutes. The MCF-7 cells were mixed with purified sheep red blood cells at a ratio of 1:10,000. Then the CTC samples were mounted onto glass slides using an 18x18 mm coverslip. Fluorescence and phase contrast image sets were acquired using a Leica DMIRE2 epifluorescence microscope equipped with a 10X objective and 12-bit monochrome CCD camera as shown in Fig.3.1((a) A phase contrast microscopy image containing two CTCs; (b) Corresponding fluorescence image shows the location of CTCs). The bright regions in fluorescence image indicate where are the CTC cells. Note that, the staining process and fluorescence imaging are used as ground truth for training and evaluating our computational algorithms. In non-invasive CTC detection, the CTC cells will not be stained so fluorescence imaging will not be used.

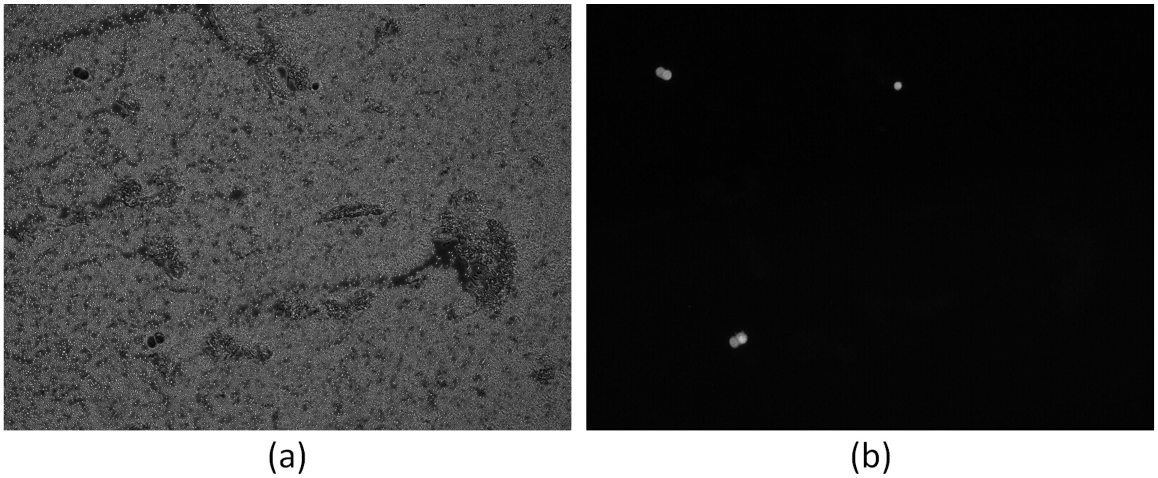


Figure 3.1. A phase contrast microscopy image and its corresponding fluorescence image.

3.2. SVM-BASED CLASSIFIER

In order to exploit the shape information of CTCs, we extract HoG features from samples in our dataset to train a SVM classifier. Since the size of CTCs varies, we normalize image patches to 64 × 64 pixels.

3.3. CNN-BASED CLASSIFIER

We adopt a CNN similar to [42] as shown in Fig.3.2. The input image patch to CNN is normalized to 40 × 40 pixels. In the first layer, 6 different convolutional filters with size 5 × 5 are applied over the input images. The convolution operation is formulated as

$$y^j = sigm(b^j + {}_{i}k^{ij} * x^i) \tag{3.1}$$

where x^i and y^j are the i-th input map and j-th output map, respectively. b^j is the bias term and k^{ij} is the convolutional kernel between x^i and y^j. The sigmoid function maps output value from -1 to 1.

The first layer is followed by a max-pooling layer which extracts local signal in every 2 × 2 region. Max-pooling function is expressed as

$$z^j_{p,q} = \max_{0 \leq m,n \leq 2} \{y^i_{2\times p+m, 2\times q+n}\} \tag{3.2}$$

where the pixel at (p, q) of the output map z^j pools over a 2 × 2 region in y^i. The third and fourth layers are another set of convolutional and max-pooling layers, and the number of kernels is 12 for both layers. The last layer is fully connected to the output layer by performing the classical dot product between their weight vector and input vector. The weighted sum is then passed to a sigmoid function.

All the parameters in kernels, bias terms and weight vectors are automatically learned by back propagation with learning rate equal to 0.1.

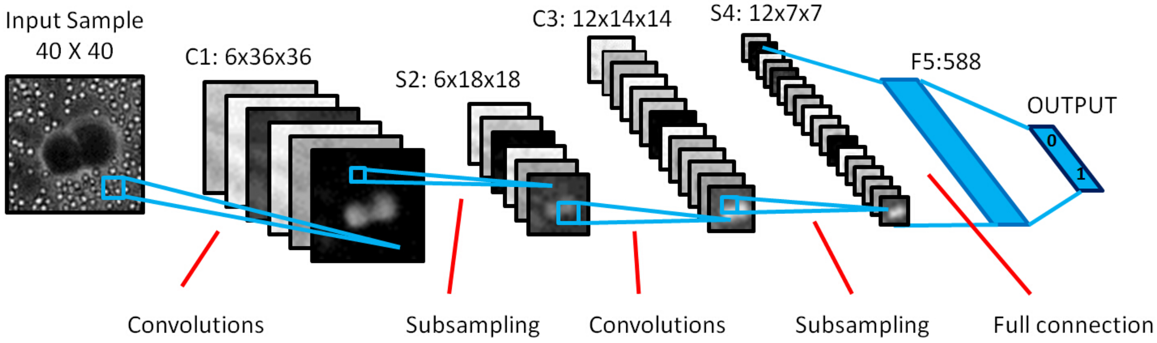


Figure 3.2. The architecture of CNN for CTC detection.

3.4. ITERATIVE TRAINING ALGORITHM

The locations of positive training samples are automatically obtained around the bright regions in fluorescence images. To consider the local context of a training sample in phase contrast microscopy image, some background pixels are cropped into the rectangular image patch as part of the positive training sample. In order to generate more positive samples for convolutional neural network and to enhance the tolerance to variations of intensity and rotation, we rotate the phase contrast microscopy images every 30 degrees and crop positive samples from them.

To find the accurate classification boundary between positive and negative samples, it is necessary to build a comprehensive negative training dataset. But collecting negative samples which cover every possible variation in the background will result in a tremendous number of negative samples, increasing the time for training. Thus, how to collect a representative set of negative samples becomes crucial.

We propose an iterative bootstrapping method to collect representative negative samples from limited training images. Compared to other training methodologies which require a large number of initial negative samples, our approach speeds up the process of training as well as refines the variation in negative samples thus improves the performance of classifiers. The proposed bootstrapping training method is summarized in Algorithm 1.

Algorithm 1 Our proposed iterative training.

```
Require:
    Initial training dataset: D;
Ensure:
    The classifier of the $i$-th iteration: $C_i$;
Initialization: $i = 0, N_{-1} = \infty$;
repeat
    Train $C_i$ on $D$;
    Perform $C_i$ on ROIs of training images.
    Gather all false alarms as $D_i$ with its number $N_i$;
    $D = D \cup D_i$;
    $i = i + 1$;
until $|N_{i-2} - N_{i-1}| < \epsilon$
```

In Algorithm 1, we define a Region-of-Interest (ROI) to reduce the search space of negative samples. An observable characteristic of CTCs is that the centers of them are always black. Negative samples around the classification boundary should share the similar features. Therefore, it is unnecessary to add samples without such kind of features to the training dataset, such as samples which are full of white blood cells. We apply a box filter on the input image (Fig.3.3(a)). Locations with low responses indicate they are dark regions, thus we consider them as negative sample regions (black in Fig.3.3(b)). There are two benefits from the ROI detection: (1) reduce the number of negative training samples, resulting in shorter training time; (2) reduce the variations in negative training samples, which makes classifiers more effective to classify hard samples.

In Algorithm 1, the initial negative samples for training are randomly cropped from ROIs of training images. In each iteration, classifiers are applied on ROIs to generate false positives. The proposed training method stops when there is no significant number of false positive samples reduced. Note that some of false positive samples in D_i may appear in the existing D. We consider these samples as important ones and we still add them to D, which increases their weights in the next training iteration to refine the decision boundaries between positive and negative samples.

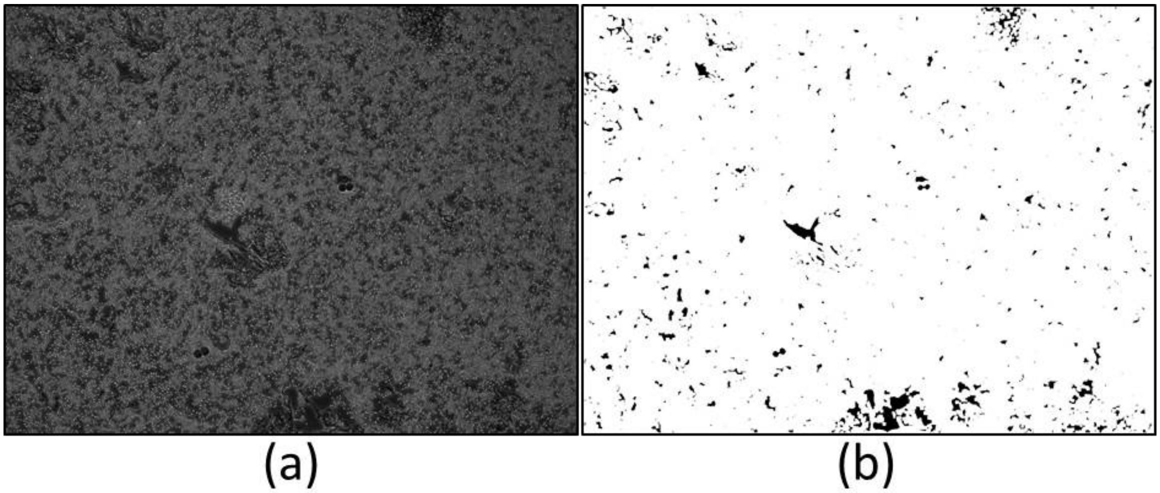


Figure 3.3. Region-of-Interest.

3.5. ROUND TRAINING ALGORITHM

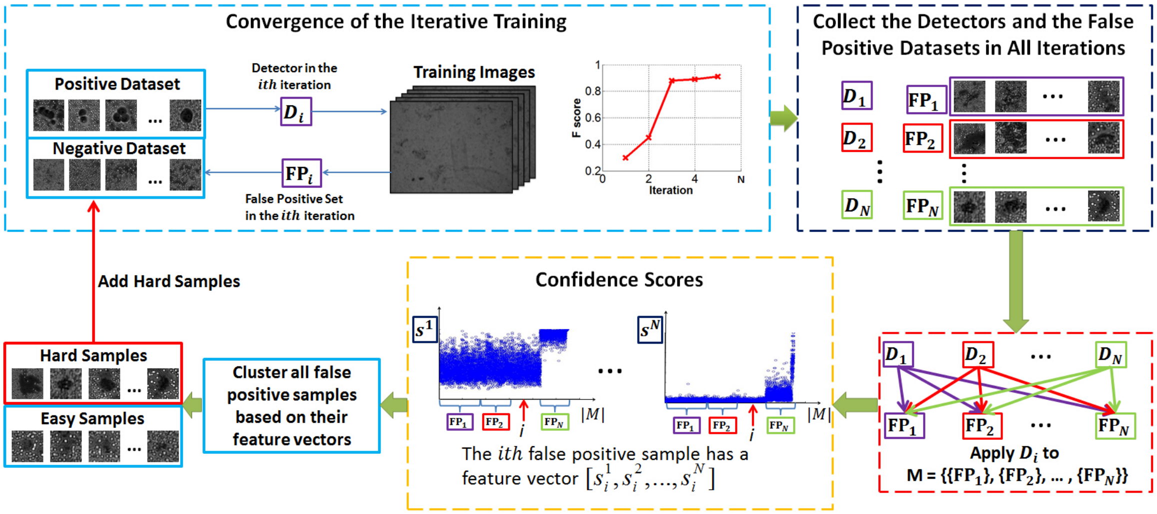
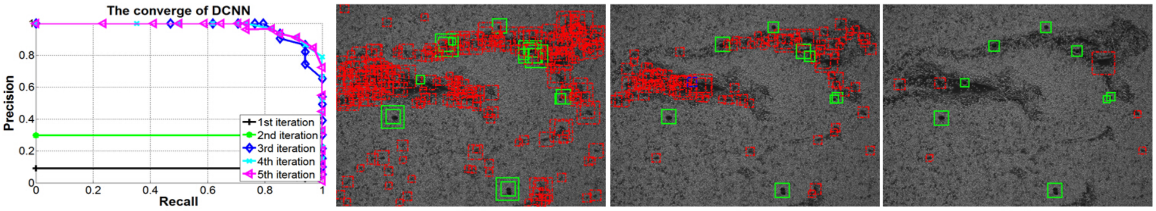

Convergence of the Iterative Training
Positive Dataset
Negative Dataset
...
Detector in the ith iteration
D_i
Training Images
FP_i
False Positive Set in the ith iteration
F score
Iteration
0
0.2
0.4
0.6
0.8
1
2
4
N
Collect the Detectors and the False Positive Datasets in All Iterations
D_1
D_2
D_N
FP_1
FP_2
FP_N
Apply D_i to M = {{FP_1}, {FP_2}, ... , {FP_N}}
Confidence Scores
s^1
s^N
$|M|$
i
The ith false positive sample has a feature vector $[s_i^1, s_i^2, ..., s_i^N]$
Cluster all false positive samples based on their feature vectors
Hard Samples
Easy Samples
Add Hard Samples

Figure 3.4. The overview of our proposed framework.

The diagram of our framework is shown in Fig. 3.4. In the ith iteration, DCNN detector D_i is trained from positive and negative training dataset plus the false positives generated from detectors D_1 to D_{i-1}, and the ith detector D_i generates a set of false positives FP_i. FP_i is added to the training dataset to train the detector D_{i+1} in the $i+1$ iteration. The iteration stops when the performance converges. This sequence of iterations is defined as one ROUND of training in the paper. Then, we apply all D_i's ($i \in [1, N]$) in one ROUND of N iterations to all the collected false positive samples during N iterations. The *confident scores* S_i are the output values of D_i to label the false positive samples as positive. Since we have N iterations in one ROUND, each false positive will have a $N \times 1$ feature vector. K-mean clustering method is applied to classify these false positives based on their feature vectors into two groups: easy samples and hard samples. Only hard samples are added to the original negative training dataset to start another ROUND of iteratively training. We obtain one DCNN detector eventually after multi-ROUNDs of training (each ROUND has multi-iterations), i.e., the final trained detector is the DCNN in the last iteration of the last ROUND.



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

Figure 3.5. Iteratively training results.

The locations of positive training samples are automatically obtained around the bright regions in fluorescence images. In order to enhance the tolerance to variations of intensity and rotation, we rotate the phase contrast microscopy images every 30 degrees and automatically crop positive samples from them.

It is important to build a comprehensive negative training dataset in order to precisely define the classification boundary between positive and negative samples. But collecting negative samples which cover every possible location in the background may introduce a lot of repetitive samples and cause a large class imbalance between positive and negative samples. Thus, how to collect a representative set of negative samples becomes crucial.

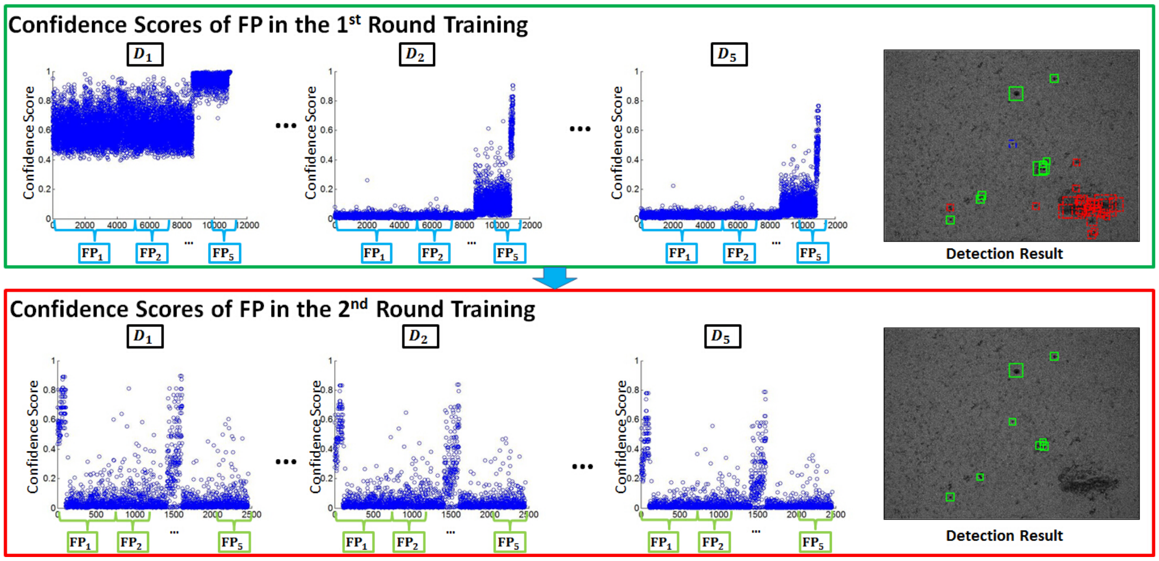


Figure 3.6. Confidence scores of false positive samples in each ROUND.

We propose a bootstrapping method to collect representative negative samples from limited training images. Unlike other training methodologies which train classifiers with all the found false positives until the performance converges, our approach continues to refine the classification boundary by training with the most representative samples among the false positives.

Traditional boosting training method trains the detector iteratively. After one iteration, the detector will collect false positives and add them to negative training dataset, and then start a new training iteration. As shown in Fig. 3.5(a), the traditional iterative training ends when the performance converges. However, the detector after the iterative training still contains quite some false positives (Fig.5(d)).

When we apply the trained detector of each iteration on all the false positives, a large amount of false positives generate low responses as shown in the first ROUND training in Fig. 3.6, which means they can be classified as negative samples relatively easily. As shown in Fig. 3.7, these relatively easy false positive samples are close to the classification

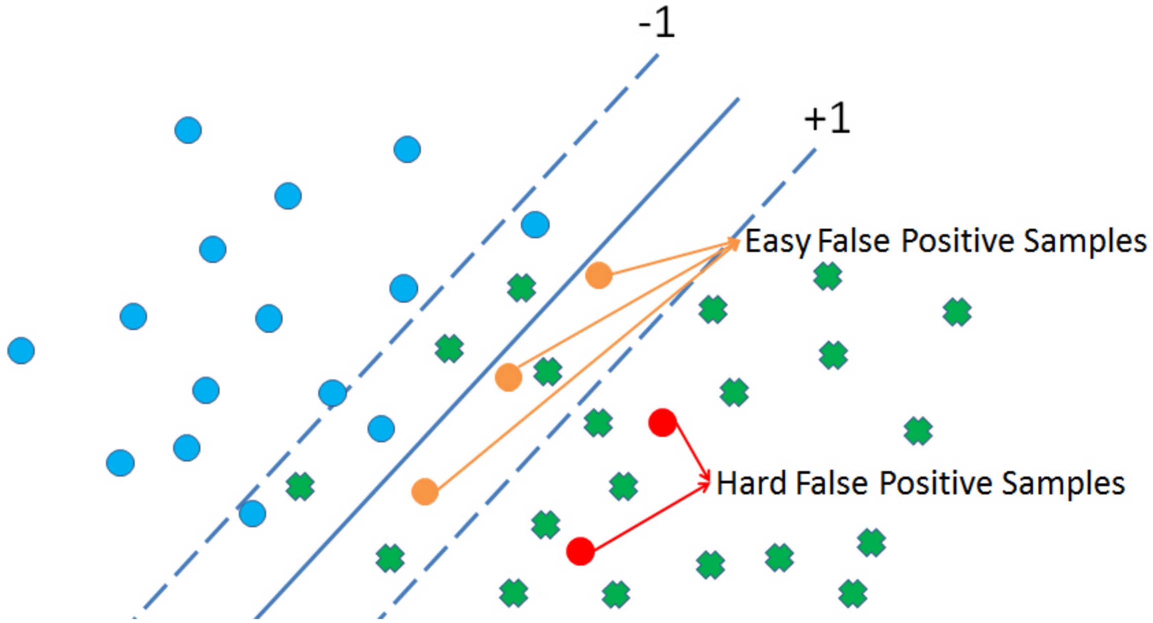


Figure 3.7. Illustration of easy and hard false positive samples.

boundary. But the rest small amount of false positive samples with relatively high responses are hard samples far away from the classification boundary. To train a classifier to better classify those hard samples, they should gain more weights in the training.

Suppose we have N iterations in one ROUND of iterative training, then we apply these N detectors on all the false positives collected from all iterations. For each false positive sample, it has a $N \times 1$ confidence score feature vector. The confidence score is the output of a classifier which indicates how likely a false positive sample is classified as positive. The higher the confidence score is, the more likely the false positive sample is classified as positive. We simply apply k-mean clustering method to classify these false positives based on confidence score feature vectors into two groups: easy samples which have low confidence scores and hard samples which have high confidence scores. To enhance the influence of these hard samples on the training, we start another ROUND of iterative training by only adding these hard samples to the previous negative training dataset. By this iterative training, only a small number of false positives will be collected. As shown in Fig. 3.6, the number of false positive samples reduces from 11000 in the first ROUND to 2500 in the second ROUND. The proportion of samples with relatively high scores in the second ROUND is larger than that in the first ROUND. Thus hard false positive samples gain more weights in the new training ROUND.

3.6. EXPERIMENTAL RESULTS

In this section, we will evaluate the performance of our proposed DCNN and the effectiveness of our training method.

3.6.1. Evaluation Metric. We acquired 45 phase contrast microscopy images, each of which has its corresponding fluorescence image as the ground truth. We randomly select 35 images for training and the rest 10 for testing. To avoid bias, we repeat this random experiment 5 times. The evaluation result is based on the average performance of 5 trials. As defined in PASCAL [43], a detection is a True Positive (TP) if the area of the intersection between the detection window and the ground truth exceeds 50 percent of their union area, otherwise it is a False Positive (FP). If one cell is not detected, it is missed (False Negative, FN). We define precision as $P = |TP|/(|TP| + |FP|)$, recall as $R = |TP|/(|TP| + |FN|)$, and F score as the Harmonic mean of precision and recall.

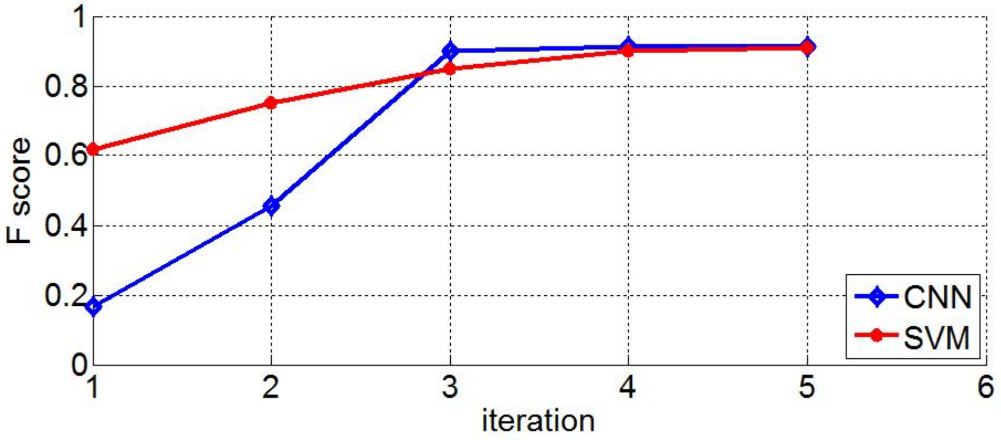


Figure 3.8. The convergence of DCNN and SVM.

Table 3.1. Evaluation on the proposed training method.

	F Score
1st Round HoG + HoC	75.4 %
1st Round DCNN	91.2 %
2nd Round HoG + HoC	78.4 %
2nd Round DCNN	97.0 %

3.6.2. Comparison of Hand-Crafted Features and Features Learned by DCNN. The Histogram-of-Gradient (HoG, [44]) feature can be used to extract regional gradient information, capturing the shape of objects. The Histogram of Color (HoC) of image patches may be considered as an feature to separate cells from the background. We distribute the color of image patches into 32 bins. In the experiment we feed the HoG + HoC to Support Vector Machine (SVM, [45]) to compare with DCNN.

The average number of positive training samples during the five trials is 1400. Training DCNN classifier takes 2 hours and training SVM takes around 0.5 hour. Note: we only use fluorescence images as ground truth. No information from fluorescence image is extracted as image features for CTC detection.

As shown in Tab. 3.1, the F score of SVM + HoG + HoC is 78.4% and that of DCNN is to 97%. The F score of DCNN is larger than that of SVM + HoG by 18.6 percentage points in the second ROUND. This result indicates that DCNN finds the better feature than HoG + HoC to detect CTCs. Some detection examples of DCNN are shown in Fig.3.9.

3.6.3. Validation of the Proposed Training Method. We evaluate our training methodology for both SVM and DCNN classifiers. Fig.3.8 shows the F score in every iteration of 2 ROUNDS. Both the SVM and DCNN classifiers converge in 5 iterations in each ROUND. The performance of both DCNN and SVM + human-designed feature improve after the first ROUND, which shows that our training method is effective in finding representative samples. Without our training method, the DCNN achieves F score of 91.2%. The F score of SVM + HoG + HoC is 75.4%. The F scores increase to 97% and 78.4% respectively with our training method, as summarized in Table 1.

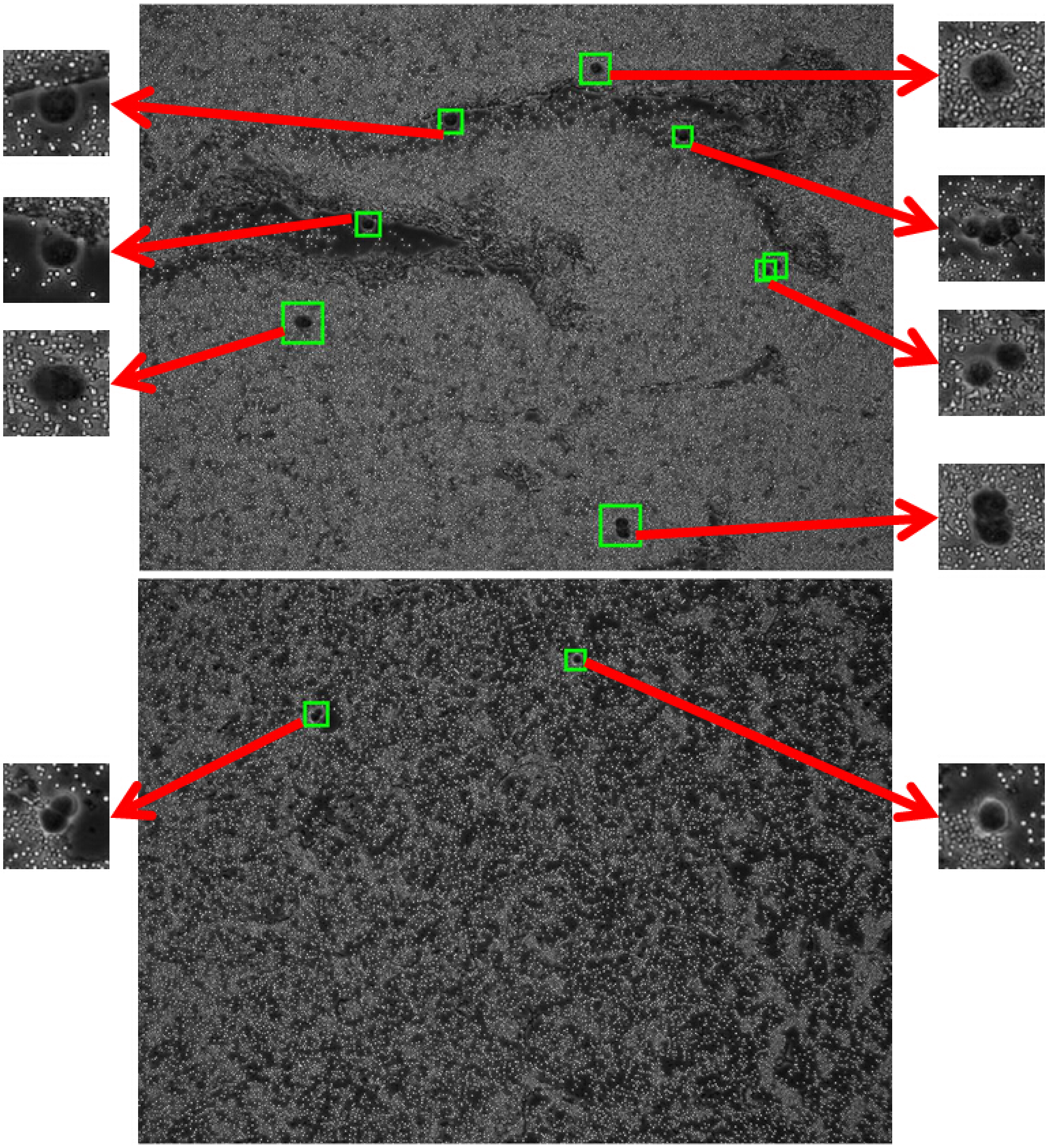


Figure 3.9. Samples of CTC detection.

4. HIERARCHICAL CONVOLUTIONAL NEURAL NETWORKS FOR MITOSIS DETECTION

Our proposed method takes a video sequence as the input, and detects when and where mitosis events occur in the sequence. It consists of two steps: first, candidate patch sequences that possibly contain mitosis events are extracted from the image sequences; then, each candidate patch sequence is classified by our Hierarchical Convolutional Neural Network (HCNN).

The first step of mitosis detection is to extract mitosis candidate sequences from the input time-lapse image sequence. The mitosis candidate extraction aims to find region-of-interest (ROI) in which are highly like to contain mitotic cells and retrieve all spatial-temporal patch sequences, while retrieving as small a number of sequences not containing mitosis patch sequences as possible. This step serves to reduce the search space. As a result, the subsequent steps can be more efficiently conducted, while maintaining mitosis detection accuracy. Fig. 4.1 shows some examples of candidate patch sequences our method automatically extracted. Our proposed mitosis candidate extraction consists of two steps: (1) salient region detection, (2) image patches retrieving.

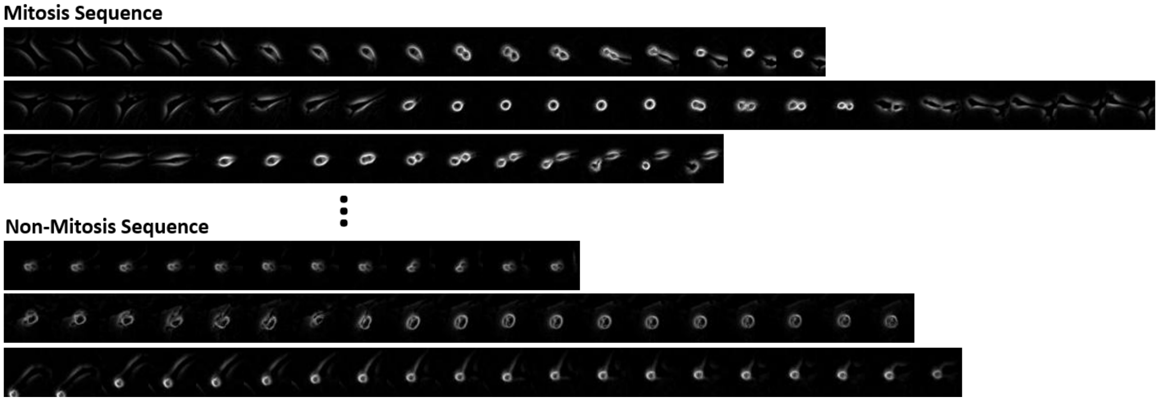


Figure 4.1. Samples of extracted candidate sequence.

4.1. SALIENT REGION DETECTION

Traditional search schemes adopt a sliding window fashion, by which the detectors need to search and classify image patches at all location. This increase the search space and potentially increase the possibility of error. Phase contrast microscopes convert the minute phase shifts caused by transparent specimens to the illuminating light source into variations in light amplitudes that can be observed by naked eyes or captured by cameras. Due to the optical principle and the inherent imperfections of the conversion process, phase contrast images contain artifacts such as halos and shade-off. If we are able to only focus on regions where mitotic cells are highly like to appear, we will be able to reduce the search space. In [34, 38], they compute the average image of original or illumination-corrected images in the given sequence, and then the average image is subtracted from each image. By this procedure, they aim to remove stationary bright artifacts. In fact, the previous procedure is only able to remove stationary artifacts. Furthermore, since the intensity values of mitotic cells are decreased when each image is subtracted from the average image, it may potentially harm the performance of later classification. The process of mitosis contain large intensity and shape change in the observed microscopy images, thus in this section we are interested in find salient regions with large intensity and shape change while maintaining the intensity values of the original images.

4.1.1. Problem Formulation. Given a time-lapse phase-contrast microscopy image data, which contains non-mitotic cells, mitotic cells, and artifacts, we are trying to find regions where are most likely to contain mitotic cells from the phase-contrast microscopy images first. This image data can be modeled as:

$$\mathcal{M} = \mathcal{L} + \mathcal{S} + \mathcal{N} \tag{4.1}$$

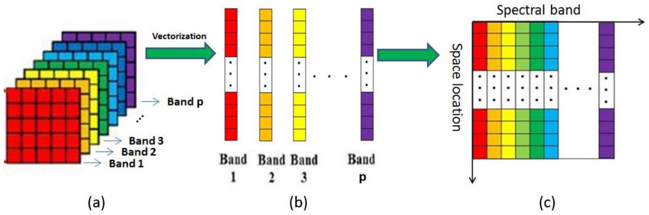


Figure 4.2. Formulation of Casorati matrix.

where $\mathcal{M} \in \mathbb{R}^{m \times n \times p}$ is the time-lapse phase-contrast microscopy image, $\mathcal{L} \in \mathbb{R}^{m \times n \times p}$ is the image containing stationary artifacts and non-mitotic cells, $\mathcal{N} \in \mathbb{R}^{m \times n \times p}$ is the Gaussian noise image, and $\mathcal{S} \in \mathbb{R}^{m \times n \times p}$ is the image containing mitosis candidates. m and n are the number of rows and columns of the microscopy image, and p is the number of images in this dataset.

We first transfer the image data matrices $\mathcal{M}$, $\mathcal{L}$, $\mathcal{S}$, and $\mathcal{N}$ to the corresponding Casorati matrices (a matrix whose each column is a vectorized image of the image data), $M \in \mathbb{R}^{mn \times p}$, $L \in \mathbb{R}^{mn \times p}$, $S \in \mathbb{R}^{mn \times p}$, and $N \in \mathbb{R}^{mn \times p}$. As shown in Fig. 4.2. Now, from Eqn.1 we have

$$M = L + S + N \tag{4.2}$$

From the phase-contrast microscopy images, we can see that only a small portion of the image contain the mitosis candidates, therefore, the matrix S is sparse, i.e., only a few elements of S are nonzero. In order to get the images which only contain mitosis candidates, we need to estimate the image S from the observed microscopy image M.

4.1.2. Low-Rank Property of Artifact Image. As shown in Fig. 4.2 (c), the Casorati matrix of the phase-contrast microscopy image data has two dimensions, the spatial domain of each image and spectral domain of all the images. From the phase-contrast microscopy images we can see that the stationary artifacts and non-mitotic cells appear at the same location of all the images. Accordingly, there exists high correlations among the spectral signatures of the artifact image data (rows of L), and each spectral signature can be represented by a linear combination of a very small number of pure spectral endmembers, which is known as the linear spectral mixing model [46] [47]. Suppose the number of pure spectral endmembers for the artifact image data L is upper bounded by r, then the rank of L is also bounded by r, i.e., $rank(L) \leqslant r$. Usually, this upper bound value of the number of endmembers r is significantly smaller than the column number and row number of L, which suggests the low-rank property of the Casorati matrix L.

Based on the low-rank property of matrix L and sparsity of matrix S, the low-rank matrix recovery (LRMR) model can be used to estimate the image L from the original phase-contrast microscopy image M.

4.1.3. LRMR Model and RPCA Problem. The low-rank matrix recovery (LRMR) model is first proposed in [48] and is considered as an idealized Robust Principal Component Analysis (RPCA) problem. For our problem, the RPCA can be formulated as follows: Given the original phase-contrast microscopy image data matrix M, the low-rank artifact and non-mitotic image data matrix L and sparse mitosis candidate matrix S are unknown, and we are trying to estimate L. This optimization problem can be formulated as

$$\min_{L,S} rank(L) + \lambda \|S\|_0 \quad s.t. \quad M = L + S \tag{4.3}$$

where $rank(\cdot)$ denotes the rank of a matrix, and λ is a positive weighting parameter. However, this is a nonconvex optimization problem, and to our best knowledge, there is no efficient solution available. A feasible solution is relaxing this problem by replacing the rank with the nuclear norm and the ℓ_0-norm with the ℓ_1-norm to obtain a tractable optimization problem [49]-[50].

$$\min_{L,S} \|L\|_* + \lambda \|S\|_1 \quad s.t. \quad M = L + S \tag{4.4}$$

The augmented Lagrangian multiplier (ALM) function of problem (4.4) is

$$\mathbf{L}(L,S,Y,\mu)=\|L\|_*+\lambda\|S\|_1+\langle Y,M-L-S\rangle+\frac{\mu}{2}\|M-L-S\|_F^2 \tag{4.5}$$

This ALM function can be solved by applying the Alternating Splitting Augmented Lagrangian Method (ASALM) [51] and the Iterative Thresholding (IT) approach [48]. More specifically, the ALM function is decomposed into two smaller subproblems which solve the variables L and S separably in the consecutive order and in an iterative way. Given $(L^{(k)},S^{(k)},Y^{(k)})$, the ASALM update the optimal solution via the following scheme until convergence:

$$\begin{cases} L^{(k+1)} = \arg\min\limits_{L} \mathbf{L}(L^{(k)},S^{(k)},Y^{(k)},\mu) \\ S^{(k+1)} = \arg\min\limits_{S} \mathbf{L}(L^{(k+1)},S^{(k)},Y^{(k)},\mu) \\ Y^{(k+1)} = Y^{(k)} + \lambda(M - L^{(k+1)} - S^{(k+1)}) \end{cases} \tag{4.6}$$

The first subproblem in (4.6) can be written into a more specific form:

$$\begin{aligned} L^{(k+1)} &= \arg\min_{L} \mathbf{L}(L^{(k)},S^{(k)},Y^{(k)},\mu) \\ &= \arg\min_{L}(\|L\|_* + \langle Y^{(k)}, M - L - S^{(k)}\rangle \\ &\quad + \frac{\mu}{2}\|M - L - S^{(k)}\|_F^2) \\ &= \arg\min_{L}(\|L\|_* + \langle Y^{(k)}, M - L - S^{(k)}\rangle \\ &\quad + \frac{\mu}{2}\|M - L - S^{(k)}\|_F^2 + \frac{(Y^{(k)})^2}{2\mu}) \\ &=\arg\min_{L}(\|L\|_*+\frac{\mu}{2}\|L-(M-S^{(k)}+\frac{Y^{(k)}}{\mu})\|_F^2) \end{aligned} \tag{4.7}$$

According to *Lemma 2.2* in [51], we can obtain the optimal solution of this function as follows:

$$L^{(k+1)} = \Psi_{\frac{1}{\mu}}(M - S^{(k)} + \frac{Y^{(k)}}{\mu}) \tag{4.8}$$

The second subproblem in (4.6) can be written into the following form:

$$\begin{aligned} S^{(k+1)} &= \arg\min_S \mathbf{L}(L^{(k+1)}, S^{(k)}, Y^{(k)}, \mu) \\ &= \arg\min_S (\|L^{(k+1)}\|_* + \langle Y^{(k)}, M - L^{(k+1)} - S\rangle \\ &\quad + \frac{\mu}{2}\|M - L^{(k+1)} - S\|_F^2) \\ &= \arg\min_S (\|L^{(k+1)}\|_* + \langle Y^{(k)}, M - L^{(k+1)} - S\rangle \\ &\quad + \frac{\mu}{2}\|M - L^{(k+1)} - S\|_F^2 + \frac{(Y^{(k)})^2}{2\mu}) \\ &= \arg\min_S (\|L^{(k+1)}\|_* + \frac{\mu}{2}\|S - (M - L^{(k+1)} + \frac{Y^{(k)}}{\mu})\|_F^2) \end{aligned} \tag{4.9}$$

According to *Lemma 2.1* in [51], we can obtain the optimal solution of this function as follows:

$$S^{(k+1)} = \Phi_{\frac{\lambda}{\mu}}(M - L^{(k+1)} + \frac{Y^{(k)}}{\mu}) \tag{4.10}$$

We summarize this method in the following form:

The k-th iteration of ASALM for problem (4.6)

Given $(L^{(k)}, S^{(k)}, Y^{(k)})$, we update them as follows:
1. $L^{(k+1)} = \Psi_{\frac{1}{\mu}}(M - S^{(k)} + \frac{Y^{(k)}}{\mu})$
2. $S^{(k+1)} = \Phi_{\frac{\lambda}{\mu}}(M - L^{(k+1)} + \frac{Y^{(k)}}{\mu})$
3. $Y^{(k+1)} = Y^{(k)} + \lambda(M - L^{(k+1)} - S^{(k+1)})$

The low-rank matrix recovery (LRMR) model is able to remove the stationary artifacts from the phase-contrast microscopy images. Moreover, as most of the cells almost stay stationary in many consequent images, they will be regarded as the low-rank component and fall into the matrix L. So a byproduct of the LRMR model is that most of the stationary cells are removed, and only the mitotic and migrating cells are picked out as the mitosis candidates, which are separated into matrix S. This can reduce the number of negative samples greatly, and as a result, the searching space and time can be decreased heavily. However, some parts of the mitotic cells may be separated into matrix L because of the absence of an appropriate spatial constraint in the LRMR model, which may cause the loss of the intensity of mitotic cells.

4.1.4. Total Variation. The total variation (TV) model has been introduced by Rudin-Osher and Fatemi in [52] as a regularization approach capable of removing noise in a given image. This model has shown great success in removing noise while at the same time significantly preserving the edge information and piecewise smooth structure. The mitosis and some abnormal cells, e.g., cells appear much brighter than normal cells, show quite different appearance from the normal cells, which can be regarded as image noise. Accordingly, the TV model can be an appropriate spatial constraint of the phase-contrast microscopy images for further extracing the edge information of the mitosis candidates. This problem can be formulated as follows:

$$L = X + P \tag{4.11}$$

After obtaining the artifacts and non-mitotic cells image L with LRMR model, we want to further separate it into two parts by means of the TV model: the final artifacts and non-mitotic image X which contains no mitosis candidates edge information, and the mitosis candidate image P which contains the edge of mitosis and some abnormal cells. Then we add P back to S to get our final mitosis candidate image.

One thing needs to be mentioned is that when we apply the LRMR model to the original phase-contrast microscopy images, the output is a Casorati matrix, i.e., each image is vectorized as a column vector, so we need to transfer the Casorati matrix $L \in \mathbb{R}^{mn \times p}$ into the normal format $\mathcal{L} \in \mathbb{R}^{m \times n \times p}$ first, then we can apply the TV model to each image separately to further seperating the mitosis candidates from the phase-contrast microscopy images. The TV model can be expressed as

$$\min_{X}\{\|X-\mathcal{L}_i\|_F^2+2\sigma TV(X)\} \tag{4.12}$$

where $\mathcal{L}_i$ is the i-th image of $\mathcal{L}$, i.e., the artifact images obtained with LRMR model, σ is a positive regularization parameter, and $TV(\cdot)$ is a discrete total variation function. For a matrix $X \in \mathbb{R}^{m\times n}$, two popular choices for the discrete TV are the isotropic TV defined by [53] [54]

$$\begin{aligned} TV_I(X) = {}^{m-1}_{i=1}\,{}^{n-1}_{j=1} &\sqrt{(X_{i,j}-X_{i+1,j})^2+(X_{i,j}-X_{i,j+1})^2} \\ &+ {}^{m-1}_{i=1}|X_{i,n}-X_{i+1,n}| + {}^{n-1}_{j=1}|X_{m,j}-X_{m,j+1}| \end{aligned} \tag{4.13}$$

and the ℓ_1-based, anisotropic TV defined by

$$\begin{aligned} TV_{\ell_1}(X) = {}^{m-1}_{i=1}\,{}^{n-1}_{j=1} &\{|X_{i,j}-X_{i+1,j}|+|X_{i,j}-X_{i,j+1}|\} \\ &+ {}^{m-1}_{i=1}|X_{i,n}-X_{i+1,n}| + {}^{n-1}_{j=1}|X_{m,j}-X_{m,j+1}| \end{aligned} \tag{4.14}$$

in this paper, the ℓ_1-based, anisotropic TV function is adopted in the TV model. The Fast iteration Shrinkage/Thresholding Algorithm (FISTA) introduced in [53] is applied to solve problem (4.12).

In different phase-contrast microscopy image data, the mitosis candidate intensity, i.e., the number of mitosis, is often different, so it is not suitable to use a constant regularization parameter σ in the TV model, which does not take this fact into consideration, to seperate the mitosis candidates with different mitosis candidate intensity. In order to overcome this problem and improve the separating performance of the TV model, we adopt an adjusted regularization parameter in our TV model to seperate images with different mitosis candidate intensity. More specifically, if the mitosis candidate intensity of an image is large, we will select a relatively big σ for the TV model, and vice versa. In our experiment, the image gradient is selected as a quantitative index of the mitosis candidate intensity in this image.

4.2. IMAGE PATCHES RETRIEVING

After image pre-processing, we are able to generate images which only contain mitosis cells, migrating cells and moving artifacts. We apply a small gaussian filter to smooth the image and threshold it into a binary mask. We calculate the area of each connected component (blob) in the binary mask. Only those blobs whose areas are above a threshold are considered as potential mitotic regions. Finally we track each connected component (blob) into candidate sequences by considering the tracking as an association problem[55], and each image patch is extracted at the fixed size $d \times d$ around the center of each connected component.

The time length of mitosis events may be quite different. However, the most salient images during the mitosis are just a few images around the birth moment, so we choose a fixed short temporal window to extract candidate patch sequences as the input to our HCNN. As for the input of TS-BLSTM, the entire patch sequence can be taken as the input.

In our experiments, the Gaussian filter has standard derivation of 3, the threshold in thresholding images is set to be 10, $d = 52$, and the minimum blob area is required to be 400 pixels in the datasets. We set all the parameters here safely to ensure that the recall of mitosis events is 100% before the classification step. The search space in the video sequence is largely reduced but the precision of mitosis events by the candidate extraction step is low (1.2%), thus we propose the HCNN in the next section to further improve the performance.

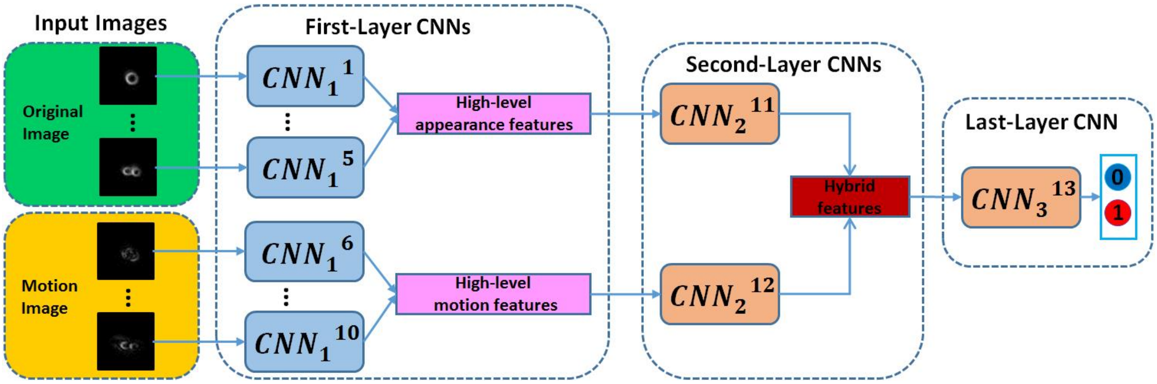


Figure 4.3. The overview of our proposed Hieratical CNN architecture

4.3. HIERARCHICAL CNN ARCHITECTURE

The overall architecture of our proposed Hierarchical Convolutional Neural Network (HCNN) is illustrated in Fig. 4.3. The first set of input contains five consecutive patches in the candidate patch sequence, and the second set of input contains the five corresponding motion images computed by the central finite difference. Each of the ten convolutional neural networks in the first layer ($CNN_1^k, k \in [1, 10]$) takes a single image as the input. In the second layer of our HCNN, we design two CNNs (CNN_2^{11} and CNN_2^{12}) to learn joint features at the patch-sequence level from original patch sequences and their motion patch sequences separately. In the last layer of our HCNN, combined appearance and motion features are fed into the last CNN (CNN_3^{13}) to make the final prediction. In the notation of CNN_i^k, i denotes the layer in our HCNN and k indexes the CNN out of the total 13 CNNs in our HCNN.

The design of such an architecture has two motivations. First, mitosis is a continuous event. Instead of detecting the mitosis events by single frame, leveraging several nearby frames will be more reliable to detect the birth moments of mitosis events. Second, the movement pattern of mitotic cells are different from that of migration cells, thus utilizing the motion information should boost the classification performance.

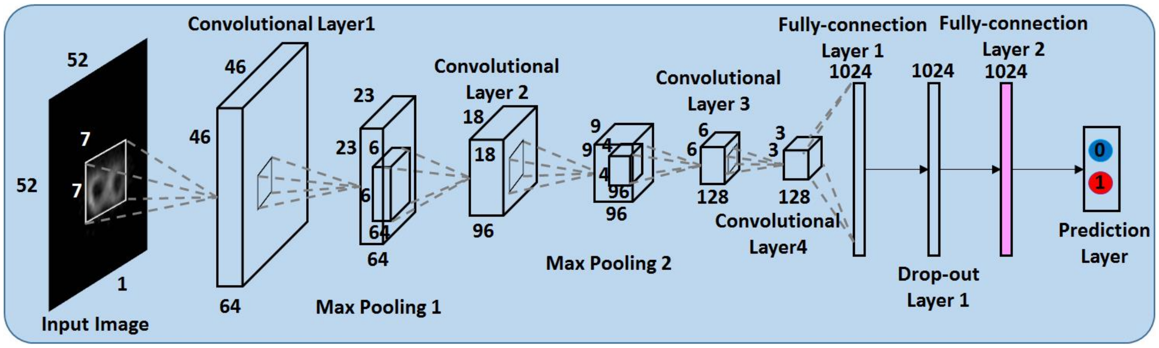


Figure 4.4. The architecture of CNNs in the first layer of our HCNN.

The first layer of our HCNN contains ten CNNs ($CNN_1^k, k \in [1,10]$), each of which classifies a single appearance or motion image at different time instants of a mitosis event. The ten CNNs shares the same architecture as shown in Fig. 5.2. There are three convolutional layers with each followed by a 2 × 2 max pooling layer. We add one more drop-out layer in case of over-fitting. The prediction layer outputs the label of the input image, indicating if the input image is the image at the specific time instant of a mitosis event.

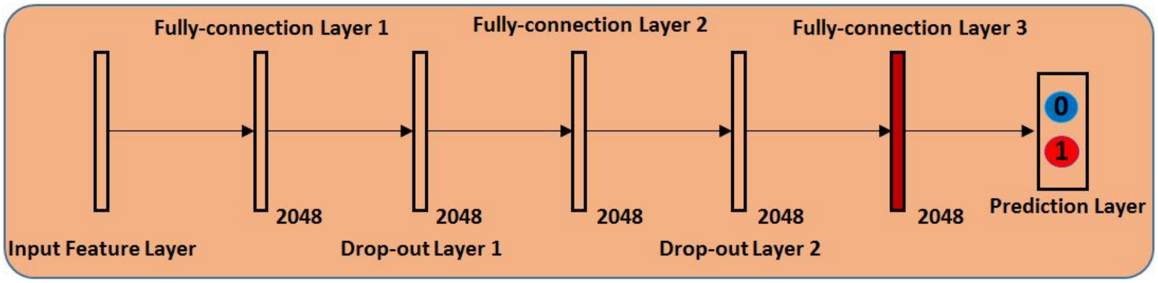


Figure 4.5. The architecture of CNNs in the second and last layer of our HCNN.

The architecture of CNNs in the second and last layer our HCNN (CNN_2^{11} CNN_2^{12} and CNN_3^{13}) is shown in Fig. 4.5. The input to CNN_2^{11} is the combined features from the Fully-connection Layer 2 of $CNN_1^k, k \in [1,5]$, leading to a 5120 vector. The input to

CNN_2^{12} is the combined features from the Fully-connection Layer 2 of $CNN_1^k, k \in [6, 10]$, and the input to CNN_3^{13} is the combined features from the Fully-connection Layer 3 of CNN_2^{11} and CNN_2^{12}.

4.4. HIERARCHICAL CNN TRAINING

Since the overall HCNN has 13 CNNs, the number of parameters is quite large. If we train the whole HCNN at once, this will increase the training complexity. Given the limited amount of training data, this will also increase the risk of over-fitting. Therefore, we divide the training process into two steps as below.

4.4.1. Pretraining Each CNN Independently. First, we train each CNN in three layers independently. The input to the first-layer CNNs ($CNN_1^k, k \in [1, 10]$) is the five original images and corresponding five motion images. For each input modality, we use the trained weights of the first CNN (e.g., CNN_1^1) as the initialization for the rest four CNNs (e.g., $CNN_1^k, k \in [2, 5]$) to achieve faster convergence. After the training on $CNN_1^k (k \in [1, 10])$ is completed, we retrieve the features from Fully-connection Layer 2 of each CNN and concatenate them as the input to the seconde-layer CNNs (CNN_2^{11} and CNN_2^{12}). The input to the third-layer CNN (CNN_3^{113}) is the concatenated features from Fully-connection Layer 3 in the second-layer CNNs. When training the 13 CNNs, we set the batch size as 100 and the number of epochs as 20 with the learning rate gradually decreasing from 10^{-2} to 10^{-4}. The drop-out rate is set to be 0.5 for all drop-out layers.

4.4.2. Fine-tuning Hieratical CNN. After each CNN is properly pretrained, we fine-tune the complete HCNN. The prediction layers of CNNs in the first and second layers are bypassed and the error from the third-layer CNN (CNN_3^{13}) is back-propagated to all the CNNs to updates the weights.

4.5. EXPERIMENTS

In the experiments, we will validate the design of HCNN and evaluate its performance.

4.5.1. Dataset. We evaluate our proposed method in five phase contrast video sequences obtained from [35], with each containing 79, 94, 85, 120 and 41 mitosis cells, respectively. Each sequence consists of 1436 images (resolution: 1392×1040 pixels). The location and time of mitosis events in the video sequences are provided as the ground truth.

In order to train our HCNN, data expansion is performed to generate more positive training data. For each positive mitosis sequence, we rotate the images every 45 degree (8 variations), slightly translate the images (e.g., by 5 pixels) horizontally and/or vertically (9 variations), which generates 72 times of the original positive training data. We retrieve negative training sequences by our proposed candidate patch sequence extraction method. At last, the training data are balanced by randomly duplicating some positive data so that the numbers of positive samples and negative samples are even.

4.5.2. Evaluation Metric. We adopt leave-one-out policy in the experiment, i.e., using four sequences for training and the rest one for testing. For testing, we use maximum-suppression to converge all the detection results based on their spatial and temporal locations and confidence scores. We use two evaluation metrics in our experiments. First, we evaluate the performance of mitosis occurrence detection in terms of the mean and standard deviation of precision, recall and F score on the five leave-one-out tests, without examining the timing of birth events. In this case, we define True Positive (TP) as a patch sequence contains a mitosis event, False Positive (FP) as it does not contain a mitosis event, and False Negative as a true positive is classified as negative. Second, the performance of mitosis detection is strictly evaluated in terms of the timing error of birth moments, i.e., those aforementioned true positive patch sequences will be considered as true positive only if the timing error of the mitosis event is equal or less than a certain threshold. The timing error is measured as the frame difference between the detection result and the ground truth.

4.5.3. Evaluation on the Hierarchical Architecture. In this section, we show the effectiveness of each module in the proposed architecture design. We compare the performance of a single-appearance CNN (CNN_1^3) targeted at the detection of the birth moment, a multi-appearance HCNN with the 5 original image patches as input (CNN_1^1 to $CNN_1^5 + CNN_2^{11}$), a simple CNN which takes 10-channel images as input and our complete HCNN. As shown in Table 5.1, because single-appearance CNN cannot capture the temporal appearance change, the F-Score of single-appearance CNN is 5 percentage points lower than that of the multi-appearance HCNN which classify the whole patch sequence. With only the appearance information as input, the F-Score of multi-appearance HCNN is 10 percentage points lower than that of our HCNN that further incorporates the motion information. As proven in [12], fusing the temporal information in feature level is better than in input pixel level, thus our HCNN performs better than a simple CNN with 10-channel images as the input.

Table 4.1. Mitosis occurrence detection accuracy of different designs.

Model	Precision (%)	Recall(%)	F score (%)
Our HCNN	99.1 ± 0.8	97.2 ± 2.4	98.2 ± 1.3
CNN with multi-channel input	97.6 ± 1.2	94.0 ± 1.9	95.8 ± 1.2
Multi-Appearance HCNN	90.9 ± 3.8	85.6 ± 3.3	88.1 ± 1.4
Single Appearance CNN	85.9 ± 4.7	80.5 ± 8.1	82.9 ± 4.7

4.5.4. Comparisons. We compare our method with six state-of-the-arts: Max-Margin Hidden Conditional Random Fields+Max-Margin Semi-Markov Model (MM-HCRF + MM-SMM) [35], EDCRF [34], HCRF [32], Hidden Markov Models (HMMs) [56], and Support Vector Machine (SVM) [57]. As shown in Table 5.2, our HCNN achieves an average precision of 99.14%, recall of 97.21 and F score of 98.15%, which outperforms the state-of-the-arts by a large margin. When evaluating the mitosis detection in term of the timing error of birth event, we use four different thresholds th (1, 3, 5 and 10) to

Table 4.2. Comparison of mitosis detection accuracy.

Model	Precision (%)	Recall(%)	F score (%)
Our HCNN	99.1 ± 0.8	97.2 ± 2.4	98.6 ± 1.3
MM-HCRF+MM-SMM	95.8 ± 1.0	88.1 ± 3.1	91.8 ± 2.0
MM-HCRF	82.8 ± 2.4	92.2 ± 2.4	87.2 ± 1.6
EDCRF	91.3 ± 4.0	87.0 ± 4.8	88.9 ± 0.7
CRF	90.5 ± 4.7	75.3 ± 9.6	81.5 ± 4.4
HMM	83.4 ± 4.9	79.4 ± 8.8	81.0 ± 3.4
SVM	68.0 ± 3.4	96.0 ± 4.2	79.5 ± 1.7

Table 4.3. Comparison of mitosis event timing accuracy.

	Precision		Recall		F score	
th	Our HCNN	[35]	Our HCNN	[35]	Our HCNN	[35]
1	92.8 ± 1.4	79.8 ± 3.4	93.1 ± 1.1	73.3 ± 2.4	93.0 ± 0.4	76.4 ± 2.7
3	96.6 ± 1.1	91.1 ± 2.2	94.9 ± 2.0	83.8 ± 3.7	95.8 ± 0.8	87.3 ±2.8
5	98.3 ± 1.2	94.7 ± 0.5	96.9 ± 1.6	87.1 ± 2.8	97.6 ± 0.9	90.8 ±1.7
10	99.1 ± 0.8	95.8 ± 1.0	97.2 ± 2.4	88.1 ± 3.1	98.2 ± 1.3	91.8 ±2.0

report the precision, recall. As shown in Table 4.3, our HCNN achieves better performance than (MM-HCRF + MM-SMM) [35]. The reason for that is two-fold. First, in [35], they extract hand-crafted SIFT features [58] from patch images, which is not the most suitable features descriptor compared with CNN; Second, their method labels each patch in the whole progress of mitosis, but the early frames and last frames may introduce noise in the model since the appearance representation of them are not clear. While we only focus on consecutive frames near the birth event, the appearance representations of these frames are clear and easy to be captured.

5. TWO-STREAM BIDIRECTIONAL LONG SHORT-TERM MEMORY FOR MITOSIS DETECTION

5.1. TWO-STREAM BIDIRECTIONAL LONG-SHORT TERM MEMORY

The overall architecture of our proposed TS-BLSTM is illustrated in Fig. 5.1. Suppose we have N appearance images $X_i, i \in [1, N]$ and their corresponding motion images M_i in one sequence. The motion images are computed simply by the frame difference. We design a CNN, as shown in Fig. 5.2, to extract the feature representation from the last fully-connected layer. Hence, appearance image X_i and motion image M_i will have feature vector f_i^x and f_i^m, respectively. Then the features of appearance images and motion images are fed into BLSTMs and generate the label l_i^x and l_i^m, respectively.

For each image, its label l_i^x predicted by appearance BLSTM and label l_i^m predicted by motion BLSTM are concatenated to make the final prediction L_i for each image in the sequence, i.e. solving the mitosis stage localization problem. To solve the mitosis detection problem, we add one more BLSTM on top of the prediction result of each image to generate the sequence label L_S. The joint objective function of our TS-BLSTM is formulated as below:

$$\min_{L_i^j, L_S}\{-T_S \log(L_S) - (1 - T_S)\log(1 - L_S) - \mathop{}_{i\in[1,N]}\,\mathop{}_{j\in[1,C]} T_i^j \log L_i^j\} \tag{5.1}$$

T_S is the label for the sequence. T_i^j and L_i^j are the label and prediction of the jth image in the ith sequence. C is the number of classes (i.e. C = 4 stages).

The two tasks we try to solve here are: (1) mitosis event detection, which is a many-to-one, binary classification problem. This requires the model to take a sequence of images as input and output one label for the whole sequence. And (2) mitosis stage localization, in which each image of one sequence is labeled to indicate which stage it

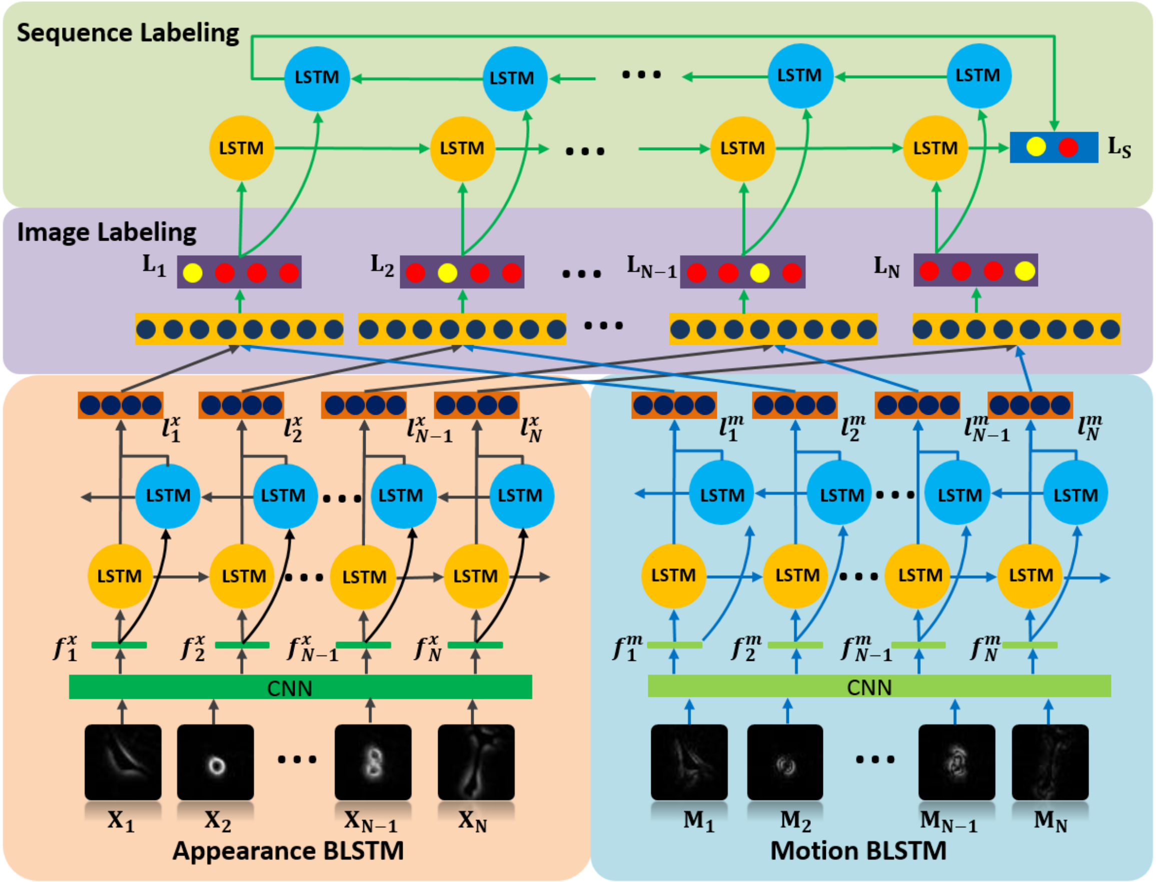


Figure 5.1. The overview of our proposed TS-BLSTM.

belongs to, can be considered as a many-to-many problem. This demands our model to be able to produce multiple types of outputs based on its multiple inputs. We unify the mitosis detection and stage localization in one architecture by combining the many-to-one model and many-to-many model in LSTMs.

Furthermore, the key to precisely label each stage in the input sequence is to locate the transition frame between two consecutive stages. When we annotated the ground truth of different stages, human experts need to look back and forth to determine which frame is exactly the transition frame between two stages. This motivates us that stage labeling should consider two directions. In our architecture, the proposed bidirectional LSTM offers the ability to unify information in both directions to label one image in the sequence.

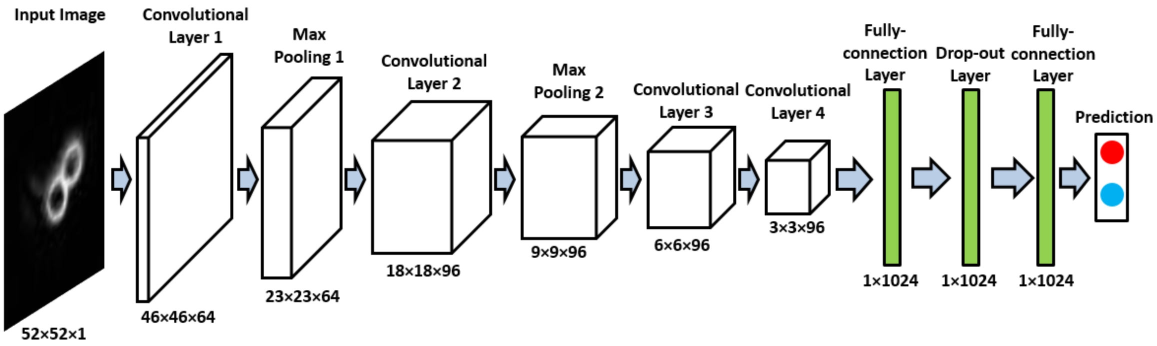


Figure 5.2. The architecture of CNN we used to extract features from the input images.

We utilize not only the appearance information, but also the motion information over time since the movement pattern of mitotic cells during different stages are different from that of migration cells. Unifying both appearance and motion cues provides rich features to describe the data thus boosts the classification performance.

When training the CNN, only the starting frame of Stage 3 is considered positive, and others are labeled as negative. Two individual CNNs are trained for the appearance input and motion input, respectively. For the training of CNN with MatConvNet, we set the patch size as 100 and the number of epoch as 20 with the learning rate gradually decreasing from 10^{-2} to 10^{-4}. The drop-out rate is set to be 0.5. When training the TS-BLSTM with Keras, we pad each training sequence to be the length of 50. The number of epoch is set as 10, and learning rate is 10^{-3} with decay rate as 10^{-6}.

5.2. EXPERIMENTS

In the experiments, we will validate the design of our proposed TS-BLSTM and evaluate its mitosis detection and stage localization performance.

5.2.1. Dataset. We use the same dataset as the HCNN. The location and time of different stages in mitosis sequences in the video are provided as the ground truth.

In order to train our CNN and TS-BLSTM, data expansion is performed to generate more positive training data to avoid the problem of overfitting. For each positive mitosis sequence, we rotate the images every 45 degree (8 variations), slightly translate the images horizontally and/or vertically (9 variations), which generates 72 times of the original positive training data. Negative sequences are extracted by the proposed candidate sequence extraction method.

5.2.2. Evaluation Metric. We adopt leave-one-out policy in the experiment, i.e., using four sequences for training and the rest one for testing. Since other competing methods classify the sequence only based on the detection of starting time of stage 3, a sequence is defined as a mitosis sequence only if it contains the starting frame of stage 3. In this case, we define True Positive as a mitosis sequence that is classified as positive, False Positive as a non-mitosis sequence that is mistakenly labeled as positive, and False Negative as a mitosis sequence that is mistakenly classified as negative.

Two evaluations are used in our experiments. First, we evaluate the performance of mitosis detection in terms of the mean and standard deviation of precision, recall and F score on the five leave-one-out tests. Second, we evaluate the performance of stage localization strictly in terms of the localization error of the starting frame of each stage. The localization error is defined as the frame difference between the detection result and the ground truth.

Table 5.1. Mitosis event detection accuracy of different designs.

Model	Precision (%)	Recall(%)	F score (%)
TS-BLSTM	98.4 ± 1.0	97.0 ± 1.8	97.7 ± 1.2
D-TS-BLSTM	94.5 ± 4.7	89.8 ± 3.8	91.8 ± 2.1
A-BLSTM	90.4 ± 6.8	95.2 ± 2.2	92.5 ± 3.2
M-BLSTM	94.4 ± 5.0	95.8 ± 4.1	95.0 ± 2.6
TS-LSTM	90.2 ± 5.3	93.0 ± 4.6	91.5 ± 2.5

5.2.3. Validation on the Proposed Architecture. In this section, we show the effectiveness of each module in the proposed architecture. We compare the performance of (1) the proposed TS-BLSTM, (2) D-TS-BLSTM (detection-only TS-BLSTM, in which only the label of sequence is predicted and the objective function does not take the classification error of each image into consideration), (3) A-BLSTM (TS-BLSTM without incorporating the motion BLSTM), (4) M-BLSTM (TS-BLSTM without incorporating the appearance BLSTM) and (5) TS-LSTM (replacing the BLSTMs in TS-BLSTM with LSTMs). As shown in 5.1, the proposed TS-BLSTM outperform other models, which shows each module (unifying mitosis event detection and stage classification, motion feature, appearance feature, and bidirectional LTSM) in the architecture of TS-BLSTM is necessary and helps boosting the performance.

Table 5.2. Comparison of mitosis event detection.

Model	Precision (%)	Recall(%)	F score (%)
Our TS-BLSTM	98.4 ± 1.0	97.0 ± 1.8	97.7 ± 1.2
HCNN	96.6 ± 1.1	94.9 ± 2.0	95.8 ± 0.8
MM-HCRF+MM-SMM	95.8 ± 1.0	88.1 ± 3.1	91.8 ± 2.0
EDCRF	91.3 ± 4.0	87.0 ± 4.8	88.9 ± 0.7
MM-HCRF	82.8 ± 2.4	92.2 ± 2.4	87.2 ± 1.6
HCRF	90.5 ± 4.7	75.3 ± 9.6	81.5 ± 4.4
HMM	83.4 ± 4.9	79.4 ± 8.8	81.0 ± 3.4
SVM	68.0 ± 3.4	96.0 ± 4.2	79.5 ± 1.7

5.2.4. Comparisons on the Mitosis Event Detection. We compare our method with seven state-of-the-arts on the performance of mitosis event detection: HCNN, Max-Margin Hidden Conditional Random Fields+Max-Margin Semi-Markov Model (MM-HCRF + MM-SMM) [35], EDCRF [34], Max-Margin Hidden Conditional Random Fields [35], HCRF [32], Hidden Markov Model (HMM) [56], and Support Vector Machine (SVM) [57]. As shown in Table 5.2, our TS-BLSTM achieves an average precision of 98.4%, recall of 97.0 and F score of 97.7%, which outperforms existing models. HCNN classifies the candidate sequence by only considering several frames nearby the starting frame of stage 3. While our model takes the whole sequence into consideration, the performance does not heavily rely on the detection of stage 3. MM-HCRF + MM-SMM [35] finishes the tasks of mitosis detection and stage localization in two separate steps, the solution cannot be jointly optimal.

Table 5.3. Comparison of stage localization accuracy.

Model	Stage 2	Stage 3	Stage 4
Our TS-BLSTM	0.78 ± 0.40	0.62 ± 0.62	0.06 ± 0.06
MM-HCRF+MM-SMM	0.82 ± 1.69	0.73 ± 1.29	1.06 ± 1.72
HCNN	N/A	0.69 ± 0.91	N/A
EDCRF	N/A	0.83 ± 1.34	N/A

5.2.5. Comparisons on the Mitosis Stage Localization. To label one mitosis sequence into the four stages, we only need to localize the starting frame of stage 2, 3, and 4. In previous work, only MM-HCRF+MM-SMM [35] is able to localize different stages while others only focus on the localization of the starting frame of stage 3. We summarize the comparison of each mitosis stage localization accuracy in Table 5.3. The results in Table 5.3 demonstrate that our method not only performs different stage localization with better performance than [35], but also achieves better accuracy for locating the starting frame of Stage 3, which is a critical point of analyzing mitosis events, than other methods.

6. CONCLUSIONS

We proposed an image-based CTC detection by two detectors: the SVM and DCNN classifiers. We also proposed an iterative training algorithm which targets at reducing the training time. To further refine the decision boundary between positive and negative samples, we proposed an effective round-based training method. Comparison of DCNN and SVM classifiers shows that our DCNN classifier works better and the proposed training method is able to improve the performance of classifiers by reducing the redundancy in negative samples. Our image-based CTC detection is not dependent on cell marker expression, and is not limited to any particular cancer type.

Further, to address the problem of mitosis event detection in phase-contrast microcopy images, we propose a Hierarchical Convolutional Neural Network (HCNN). We extract candidate patch sequences from the image sequence as the input to HCNN. In our HCNN architecture, we utilize both the appearance information and temporal cues hidden in patch sequences to identify the birth event of mitotic cells. Given the complex HCNN structure, we propose an efficient training methodology to learn the parameters inside HCNN and prevent the risk of over-fitting. In the experiments, we prove that the design of our HCNN is sound and our method outperforms other state-of-the-art by a large margin.

Considering the drawbacks of HCNN, further we propose a Two-Stream Bidirectional Long Short-Term Memory (TS-BLSTM) to tackle the two problems of mitosis event detection and stage localization jointly in phase-contrast microscopy images. Both appearance and motion information are utilized to provide rich feature description. Bidirectional LSTM helps to utilize information in both directions. In the experiments, we validate the proposed architecture and that our model outperforms other state-of-the-arts in both two tasks.

REFERENCES

[1] Yann LeCun. Gradient-based learning applied to document recognition. 1998.

[2] Yoshua Bengio, Aaron C. Courville, and Pascal Vincent. Representation learning: A review and new perspectives. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35:1798–1828, 2013.

[3] Dinggang Shen, Guorong Wu, and Heung-Il Suk. Deep learning in medical image analysis. *Annual review of biomedical engineering*, 19:221–248, 2017.

[4] Geert J. S. Litjens, Thijs Kooi, Babak Ehteshami Bejnordi, Arnaud Arindra Adiyoso Setio, Francesco Ciompi, Mohsen Ghafoorian, Jeroen van der Laak, Bram van Ginneken, and Clara I. Sánchez. A survey on deep learning in medical image analysis. *Medical image analysis*, 42:60–88, 2017.

[5] Emilian Racila, David Michael Euhus, Aaron Josef Weiss, C L Rao, JoAnn McConnell, Leon W. M. M. Terstappen, and Jonathan W . Uhr. Detection and characterization of carcinoma cells in the blood. *Proceedings of the National Academy of Sciences of the United States of America*, 95 8:4589–94, 1998.

[6] Tanja N. Fehm, Arthur I. Sagalowsky, Edward J. Clifford, Peter D. Beitsch, Hossein M Saboorian, David Michael Euhus, Songdong Meng, Larry E Morrison, Thomas F. Tucker, Nancy L Lane, B. Michael Ghadimi, Kerstin Heselmeyer-Haddad, Thomas Ried, Chandra Sekhara Rao, and Jonathan W . Uhr. Cytogenetic evidence that circulating epithelial cells in patients with carcinoma are malignant. *Clinical cancer research : an official journal of the American Association for Cancer Research*, 8 7:2073–84, 2002.

[7] W Jeffery Allard, Jeri Matera, Michael C. Miller, Madeline I Repollet, Mark C. Connelly, C L Rao, Alison T Stopeck, and L V M M Terstappen. Tumor cells circulate in the peripheral blood of all major carcinomas but not in healthy subjects or patients with non-malignant diseases. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*, 22 14_suppl:9552, 2004.

[8] Klaus Pantel, Ruud H. Brakenhoff, and Burkhard H. Brandt. Detection, clinical relevance and specific biological properties of disseminating tumour cells. *Nature Reviews Cancer*, 8:329–340, 2008.

[9] Kang Li, Mei Chen, and Takeo Kanade. Cell population tracking and lineage construction with spatiotemporal context. *Medical image analysis*, 12 5:546–66, 2007.

[10] Carl-Magnus Svensson, Solveigh Krusekopf, and Jörg Lücke and Marc Thilo Figge. Automated detection of circulating tumor cells with naive bayesian classifiers. *Cytometry. Part A : the journal of the International Society for Analytical Cytology*, 85 6:501–11, 2014.

[11] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton. Imagenet classification with deep convolutional neural networks. *Commun. ACM*, 60:84–90, 2012.

[12] Andrej Karpathy, George Toderici, Sanketh Shetty, Thomas Leung, Rahul Sukthankar, and Fei fei Li. Large-scale video classification with convolutional neural networks. *2014 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1725–1732, 2014.

[13] Dan C. Ciresan, Alessandro Giusti, Luca Maria Gambardella, and Jürgen Schmidhuber. Mitosis detection in breast cancer histology images with deep neural networks. *Medical image computing and computer-assisted intervention : MICCAI ... International Conference on Medical Image Computing and Computer-Assisted Intervention*, 16 Pt 2:411–8, 2013.

[14] Mehdi Habibzadeh, Adam Krzyzak, and Thomas Fevens. White blood cell differential counts using convolutional neural networks for low resolution images. In *ICAISC*, 2013.

[15] Hang Su, Subhransu Maji, Evangelos Kalogerakis, and Erik G. Learned-Miller. Multi-view convolutional neural networks for 3d shape recognition. *2015 IEEE International Conference on Computer Vision (ICCV)*, pages 945–953, 2015.

[16] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural computation*, 9 8:1735–80, 1997.

[17] Jeff Donahue, Lisa Anne Hendricks, Sergio Guadarrama, Marcus Rohrbach, Subhashini Venugopalan, Kate Saenko, and Trevor Darrell. Long-term recurrent convolutional networks for visual recognition and description. *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2625–2634, 2015.

[18] Philip Th Went, Alessandro Lugli, Sandra Meier, Marcel Bundi, Martina Mirlacher, Guido Sauter, and Stephan R Dirnhofer. Frequent epcam protein expression in human carcinomas. *Human pathology*, 35 1:122–8, 2004.

[19] Robert Koenigsberg, Eva Obermayr, Giovanna Bises, Georg H Pfeiler, Margit Gneist, Fritz Wrba, Maria Carmen De Santis, Robert Zeillinger, Marcus Hudec, and Christian Dittrich. Detection of epcam positive and negative circulating tumor cells in metastatic breast cancer patients. *Acta oncologica*, 50 5:700–10, 2011.

[20] Udo Bilkenroth, Helge Taubert, Dagmar Riemann, U. Rebmann, Hans Heynemann, and Axel Meye. Detection and enrichment of disseminated renal carcinoma cells from peripheral blood by immunomagnetic cell separation. *International journal of cancer*, 92 4:577–82, 2001.

[21] Hadi Esmaeilsabzali, Timothy V. Beischlag, Michael E. Cox, Ash Parameswaran, and Edward J. Park. Detection and isolation of circulating tumor cells: principles and methods. *Biotechnology advances*, 31 7:1063–84, 2013.

[22] Olivier Debeir, Philippe Van Ham, Robert Kiss, and Christine Decaestecker. Tracking of migrating cells under phase-contrast video microscopy with combined mean-shift processes. *IEEE transactions on medical imaging*, 24 6:697–711, 2005.

[23] Dirk R. Padfield, Jens Rittscher, Nick Thomas, and Badrinath Roysam. Spatio-temporal cell cycle phase analysis using level sets and fast marching methods. *Medical image analysis*, 13 1:143–55, 2009.

[24] Omar Al-Kofahi, Richard J. Radke, Susan Goderie, Qin Shen, Sally Temple, and Badrinath Roysam. Automated cell lineage construction: a rapid method to analyze clonal development established with murine neural progenitor cells. *Cell cycle*, 5 3:327–35, 2006.

[25] Kang Li, Eric D. Miller, Mei Chen, Takeo Kanade, Lee E. Weiss, and Phil G. Campbell. Computer vision tracking of stemness. In *ISBI*, 2008.

[26] Lichen Liang, Xiaobo Zhou, Fuhai Li, S.T.C. Wong, Jeremy Huckins, and R. King. Mitosis cell identification with conditional random fields. *2007 IEEE/NIH Life Science Systems and Applications Workshop*, pages 9–12, 2007.

[27] John D. Lafferty, Andrew McCallum, and Fernando Pereira. Conditional random fields: Probabilistic models for segmenting and labeling sequence data. In *ICML*, 2001.

[28] Chao-Hui Huang and Hwee-Kuan Lee. Automated mitosis detection based on exclusive independent component analysis. In *ICPR*, 2012.

[29] Greg M. Gallardo, Fuxing Yang, Fiorenza Ianzini, Michael Mackey, and Milan Sonka. Mitotic cell recognition with hidden markov models. 5367:661–668, 05 2004.

[30] Jianxin Wu, S. Charles Brubaker, Matthew D. Mullin, and James M. Rehg. Fast asymmetric learning for cascade face detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30:369–382, 2008.

[31] Paul A. Viola and Michael J. Jones. Robust real-time object detection. 2001.

[32] A.-A. Liu, K. Li, and T. Kanade. Mitosis sequence detection using hidden conditional random fields. In *ISBI*, 2010.

[33] Ariadna Quattoni, Sy Bor Wang, Louis-Philippe Morency, Michael Collins, and Trevor Darrell. Hidden conditional random fields. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29, 2007.

[34] Seungil Huh, Dai Fei Elmer Ker, Ryoma Bise, Mei Chen, and Takeo Kanade. Automated mitosis detection of stem cell populations in phase-contrast microscopy images. *IEEE transactions on medical imaging*, 30 3:586–96, 2011.

[35] Anan Liu, Kang Li, and Takeo Kanade. A semi-markov model for mitosis segmentation in time-lapse phase contrast microscopy image sequences of stem cell populations. *IEEE transactions on medical imaging*, 31 2:359–69, 2012.

[36] Zhicheng Yan, Hao Zhang, Robinson Piramuthu, Vignesh Jagadeesh, Dennis DeCoste, Wei Di, and Yizhou Yu. Hd-cnn: Hierarchical deep convolutional neural network for large scale visual recognition. 2015.

[37] Yi Sun, Xiaogang Wang, and Xiaoou Tang. Hybrid deep learning for face verification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 38:1997–2009, 2013.

[38] Yunxiang Mao and Zhaozheng Yin. A hierarchical convolutional neural network for mitosis detection in phase-contrast microscopy images. In *MICCAI*, 2016.

[39] Weizhi Nie, Wenhui Li, Anan Liu, Tong Hao, and Yuting Su. 3d convolutional networks-based mitotic event detection in time-lapse phase contrast microscopy image sequences of stem cell populations. *2016 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 1359–1366, 2016.

[40] Shuiwang Ji, Wei Xu, Ming Yang, and Kai Yu. 3d convolutional neural networks for human action recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35:221–231, 2010.

[41] Du Tran, Lubomir D. Bourdev, Rob Fergus, Lorenzo Torresani, and Manohar Paluri. Learning spatiotemporal features with 3d convolutional networks. *2015 IEEE International Conference on Computer Vision (ICCV)*, pages 4489–4497, 2015.

[42] Ting Chen and Christophe Chefd'Hotel. Deep learning based automatic immune cell detection for immunohistochemistry images. In *MLMI*, 2014.

[43] Mark Everingham, Luc Van Gool, Christopher K. I. Williams, John M. Winn, and Andrew Zisserman. The pascal visual object classes (voc) challenge. *International Journal of Computer Vision*, 88:303–338, 2009.

[44] Navneet Dalal and Bill Triggs. Histograms of oriented gradients for human detection. *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, 1:886–893 vol. 1, 2005.

[45] Corinna Cortes and Vladimir Vapnik. Support-vector networks. *Machine Learning*, 20:273–297, 1995.

[46] Mohammad Golbabaee and Pierre Vandergheynst. Hyperspectral image compressed sensing via low-rank and joint-sparse matrix recovery. *2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 2741–2744, 2012.

[47] Marian-Daniel Iordache, José M. Bioucas-Dias, and Antonio J. Plaza. Sparse unmixing of hyperspectral data. *IEEE Transactions on Geoscience and Remote Sensing*, 49:2014–2039, 2011.

[48] John Wright, Arvind Ganesh, Shankar R. Rao, YiGang Peng, and Yi Ma. Robust principal component analysis: Exact recovery of corrupted low-rank matrices via convex optimization. In *NIPS*, 2009.

[49] Emmanuel J. Candès and Benjamin Recht. Exact matrix completion via convex optimization. *Foundations of Computational Mathematics*, 9:717–772, 2009.

[50] YiGang Peng, Arvind Ganesh, John Wright, Wenli Xu, and Yi Ma. Rasl: Robust alignment by sparse and low-rank decomposition for linearly correlated images. *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 763–770, 2010.

[51] Min Tao and Xiaoming Yuan. Recovering low-rank and sparse components of matrices from incomplete and noisy observations. *SIAM Journal on Optimization*, 21:57–81, 2011.

[52] Nonlinear total variation based noise removal algorithms.

[53] Amir Beck and Marc Teboulle. Fast gradient-based algorithms for constrained total variation image denoising and deblurring problems. *IEEE Transactions on Image Processing*, 18:2419–2434, 2009.

[54] Antonin Chambolle. An algorithm for total variation minimization and applications. *Journal of Mathematical Imaging and Vision*, 20:89–97, 2004.

[55] Yunxiang Mao, Haohan Li, and Zhaozheng Yin. Who missed the class? - unifying multi-face detection, tracking and recognition in videos. In *ICME*, 2014.

[56] Lawrence R. Rabiner. A tutorial on hidden markov models and selected applications in speech recognition. 1989.

[57] Johan A. K. Suykens and Joos Vandewalle. Least squares support vector machine classifiers. *Neural Processing Letters*, 9:293–300, 1999.

[58] David G. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60:91–110, 2004.

VITA

Yunxiang Mao was born in Chengdu, China. In July 2012, he received his Bachelor's degree of Engineering in University of Electronic Science and Technology of China. Then, he joined the Computer Science Department of Missouri University of Science and Technology (formerly the University of Missouri-Rolla) as a Ph.D. student in Aug, 2012. In May, 2018, he received his Ph.D. degree in the Computer Science Department of Missouri University of Science and Technology.