8-25-2021 2:30 PM

Mitosis Detection from Pathology Images

Jinhang Zhang, *The University of Western Ontario*

Supervisor: Charles Ling, *The University of Western Ontario*

A thesis submitted in partial fulfillment of the requirements for the Master of Science degree in Computer Science

Abstract

In the case of breast cancer, according to the Nottingham Grading System, counting mitotic cells is an important indicator of tumour diagnosis and grading. Pathologists usually manually count mitosis from histopathology images to determine the cancer grade. This is a challenging and time-consuming procedure. In most recent works, different deep neural networks have been designed to detect the suspicious cells initially and count the number of them afterwards. However, these detection approaches have certain limitations including complicated structures, the detection performance is still not satisfactory, and the need of a large number of labeled images to train a satisfied model. In this paper, we modify and improve a popular one-stage object-detection deep network to facilitate the mitotic cells detection task. Our novel improvements include using different loss functions for cells of different sizes, utilizing new data augmentation methods, generating prior anchor boxes with approximate sizes, and so on. We validate our deep learning model on two public benchmark datasets named Mitosis Detection in Breast Cancer Histological Images (MITOSIS). The experimental results indicate that our method achieves the competitive results on MITOSIS-2012 dataset and on the MITOSIS-2014 dataset with faster inference speed. More importantly, we design an interactive system with "correction and relearning" pipeline so that our system can relearn from a small number of slides from a new lab and achieve satisfactory results. We design a web portal (http://ai4path.ca/#/) where this online pipeline can be easily utilized by pathologists in Western Hospital Pathology Group(WHPG) and hopefully in the future, by all pathologists in the world.

Keywords: Mitosis detection, Computer Vision, Correction-and-Relearning Pipeline.

Summary for Lay Audience

Mitotic cells detection is an important step in the pathology domain. It is mainly used to diagnose and prognose cancer in different regions of our body, for instance, breast, glioma, and melanoma etc. However, most of these examinations by pathologists under microscopes are still manual work which is time-consuming, challenging and subjective. Therefore, it is significant to provide a real-time computational examination tool for pathologists. In recent years, with the development of Convolutional Neural Network (CNN) in the computer vision domain, computers can have the ability to detect objects in images. The success in the field of computer vision has attracted the attention of researchers in the pathology domain. Previous attempts are developed based on constructing manual features and obtaining features from deep neural networks. Both of these approaches are widely deployed to accomplish the mitotic cells detection task and achieve certain results. However, the detection performance of these current methods are still not satisfactory.

In this thesis, we primarily deal with the mitotic cells detection task through modifying and improving a popular object detection method. In addition, we extract the learned knowledge from previous work as prior information. We integrate the prior information in our model to create a robust real-time tool for pathologists to examine the detected results.

Acknowlegements

To my supervisor, Dr. Charles Ling, I owe an immense debt of gratitude for his support, mentorship, scientific insights and contagious enthusiasm during my studies. His consistent, patient confidence in me was essential in the performance of the work described in this thesis. He provides much help since I came to Canada, I believe he considerably exceeded the expectations of his role as supervisor.

I would like to thank Yan Tong, YuanYuan Han, Juncheng Yin, Dr. Zhang for their encouragement and advices during this research and the life in Canada.

I would also like to show special gratitude to my parents for their constant support throughout my research and thesis writing.

Contents

List of Figures

List of Tables

List of Appendices

Chapter 1

Introduction

In this chapter, we describe our motivation and some background information on the mitosis detection task. We also discuss the difficulties and our contributions toward solving them. Furthermore, we introduce our "correction and relearning" pipeline. Finally, the thesis layout is presented at the end.

1.1 Description of the problem

Breast cancer has been the most prevalent cancer diagnosed in women. However, it can be treated successfully if recognized early. Treatment strategies should be based on the grade and prognosis of the disease. The Nottingham Grading System is widely used to determine the grade of breast cancer, and considers three essential morphological features: mitotic cells count, tubule information, and nuclear pleomorphism, among which, the mitotic cells count is the most important [23]. In practice, pathologists usually manually detect mitosis in the histopathological slides of the breast. However, these manually examinations by pathologists are time-consuming, challenging and subjective. Therefore, it is necessary to develop an automatic detection method to improve the reliability of pathological examinations and to save resources.

Mitosis detection is an common and important task in the pathology domain. It is mainly

used to diagnose and prognose cancer in different regions of our body, for instance, breast, glioma, and melanoma etc. However, it is still difficult to develop an automatic, computational examination tool for pathologists as it comprises several challenges. Firstly, accurately annotated images are required for training. However, it is time-consuming to annotate cells in each image which limits the size of the training dataset. Secondly, the morphology features of mitosis are diverse. In particular, there are four main phases of mitosis including prophase, metaphase, anaphase and telophase. As shown in the figure 1.1. In addition, the morphology of some apoptotic non-mitosis look similar to that of mitosis. This complicates the accurate mitosis detection. As shown in the figure 1.2.

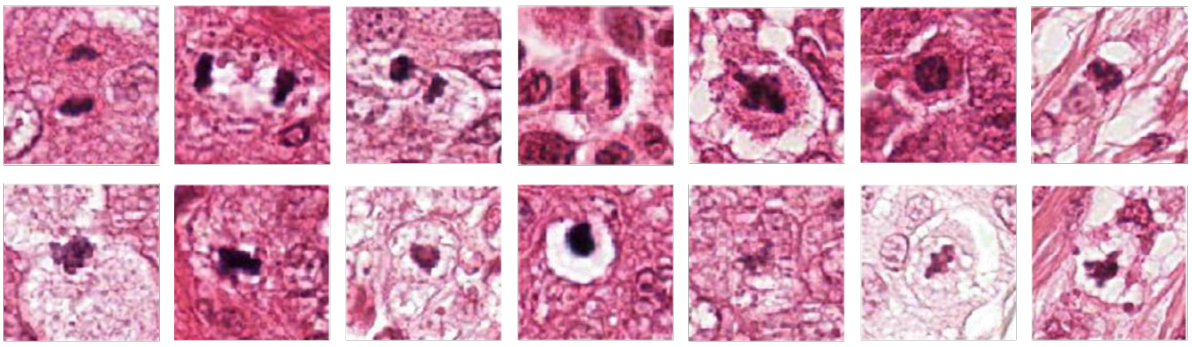


Figure 1.1: Examples of mitosis[23].

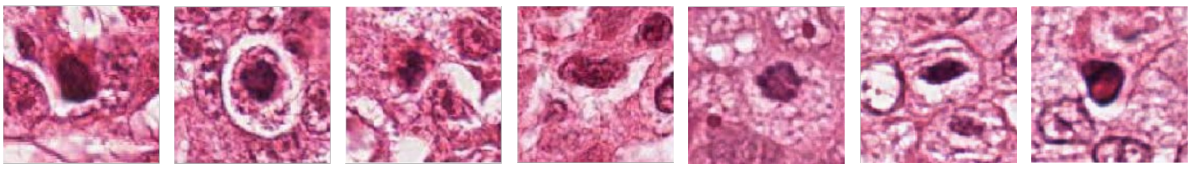


Figure 1.2: Examples of non-mitosis[23].

Thirdly, in many cases, the sizes of some mitosis are small. These cells also usually appear sparsely in a large and high-resolution slide. As shown in the figure 1.3 with annotations.

Mitosis detection is a difficult and challenging task. Previous scholars have used artificial intelligence (AI) techniques to automatically detect the mitosis. These works are mainly developed and evaluated on two widely used dataset named International Conference on Pattern Recognition (ICPR) - Mitosis Detection in Breast Cancer Histological Images (MITOSIS) in 2012 and 2014. These competition datasets contains annotated pathological slides from breasts,

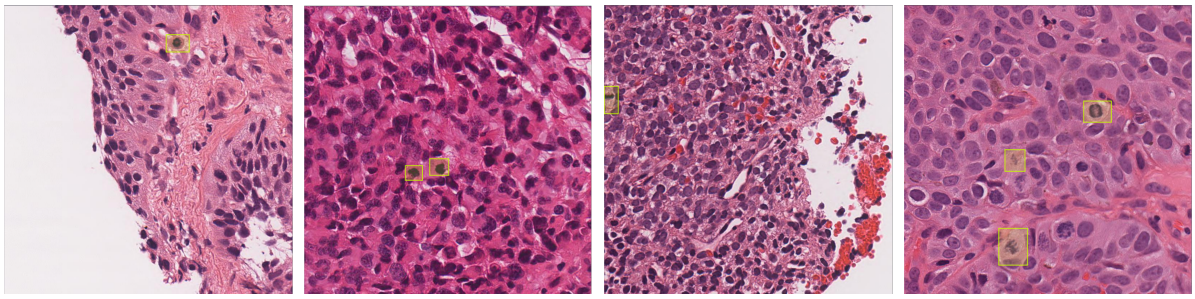


Figure 1.3: Examples of sparse cells.

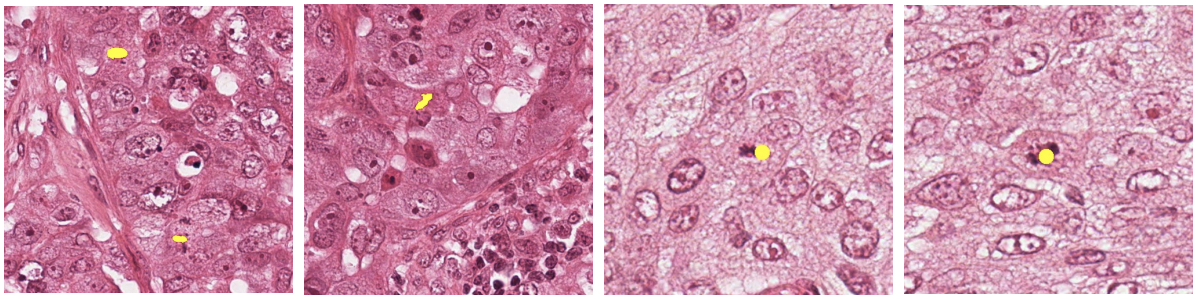


Figure 1.4: Examples of ICPR MITOSIS 2012 and 2014.

have been verified by different experts, and aims to promote the development of mitosis detection. Several examples from the ICPR MITOSIS 2012 and ICPR MITOSIS 2014 dataset are demonstrated in figure 1.4. Mitosis are annotated in pixels in the ICPR MITOSIS 2012 dataset, however, only the centroids of mitosis are annotated in ICPR MITOSIS 2014 dataset.

These previous work can be mainly divided into two categories. The first one is extracting features such as color, morphology, and texture from the region-of-interests (ROIs) by conventional image-processing methods [19, 47, 44, 34, 14]. These features are then fed into a machine-learning based classifier to determine whether the object in the ROI is a mitosis or not. This approach demonstrate good performance. However, it is expensive to construct the features. Instead of manually creating features, many researchers have adopted deep learning based methods to extract abstract features from the ROIs and to detect mitosis [6, 48, 4, 23, 22, 31]. These methods are developed based on some two-stage object detection frameworks with additional networks for post-processing. However, the networks utilized in these works can not effectively extract significant features, and these two-stage detection methods need to explicitly generate region proposals via a specific neural network on a pathological

slide which is relatively large. These two reasons lead to dissatisfying detection performance. With the development in computer vision, some one-stage detection methods can generate more accurate results and infer faster than the two-stage methods on open source datasets such as MS COCO [1] and PASCAL VOC dataset [2], owing to the use of powerful networks in the extraction of features as well as not having to explicitly generate region proposals.

In order to solve the aforementioned difficulties and improve the detection performance, in this thesis, we modify a popular one-stage object-detection framework named YOLOv4 [1] and compare the modified model with other methods proposed on the ICPR MITOSIS-2012 and ICPR MITOSIS-2014 datasets. The experiment results illustrates that our method achieves competitive results on the ICPR MITOSIS 2012 dataset and on the ICPR MITOSIS 2014 dataset with faster inference speed.

In a practical scenario, it is difficult to obtain a large number of images with enough annotated mitosis. In general, a limited number of annotated samples will lead to a poor performance of detection model. Some mitosis are not detected, and some other cells are detected. Therefore, it is critical to have an approach which can perform a good detection performance on mitosis from only a few number of images. We have collaborated with the pathologists at Western Hospital in developing a real-time detection tool. This tool aims to automatically diagnose diseases by detecting mitosis in a small number of pathological slides. In addition, we integrate an online progressive "correction and relearning" pipeline into this tool for pathologists to correct the wrong detections by re-annotating the cells. Afterwards, a new model will be trained based on these accumulative samples. Figure 1.3 and 1.5 show several examples from the Western Hospital Pathology Group (WHPG) dataset with annotations. This dataset is mainly cropped from glioma, melanoma, and meningioma, which is annotated by two pathologists. Satisfactory results can been achieved through annotating (and re-annotating) only a small number of slides by the pathologists in the WHPG. Furthermore, a web portal (http://ai4path.ca/#/) is designed so that this online pipeline can be easily utilized by the

[1] MS COCO
[2] PASCAL VOC

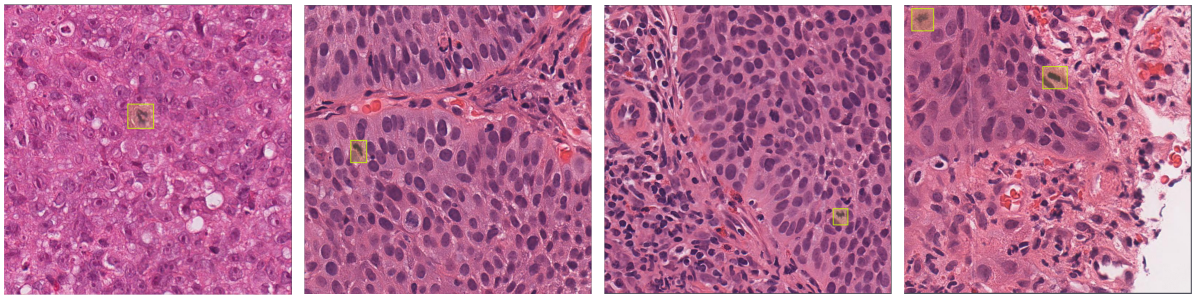


Figure 1.5: Examples of WHPG.

pathologists in the WHPG and hopefully by all of the pathologists in the world in the future. Through this application, it has been shown that our model can be used to address the real-time practical mitosis detection problem.

1.2 Contributions

In this thesis, we modify a popular one-stage object detection method named YOLOv4 to accomplish the mitosis detection task on the ICPR MITOSIS 2012 and ICPR MITOSIS 2014 dataset. Our proposed model is lightweight after the modification of the number of prediction heads. We generate prior anchor boxes with approximate sizes based on the training dataset. Also, a different method is utilized to suppress the redundant detection boxes, and different loss functions are integrated for cells with different sizes. In addition, the copy-and-paste augmentation methods are deployed to increase the occurrences of the target cells in the images. The experiment results illustrate that our method achieves competitive results of 76% precision, 88% recall, and 81.6% F1-score on ICPR MITOSIS 2012 dataset and 54% precision, 60% recall, and 56.8% F1-score on the ICPR MITOSIS 2014 dataset with faster inference speed.

Furthermore, we develop an online progressive "correction and relearning" pipeline for pathologists to correct wrong detections by re-annotating the cells. This pipeline aims to progressively learn from a few number of images. It is illustrated that the detection performance is boosted with the increased number of images corrected by pathologists through our pipeline.

1.3 Overview of the Dissertation

This thesis aims to improve the performance in mitosis detection tasks and develop a computational tool for automatic disease diagnosis based on a small number of histopathological slides. In Chapter 1, the research problem, previous researches and our contributions are introduced briefly. In Chapter 2, the background knowledge about previous works on the mitosis detection, some object detection methods, and the copy-and-paste augmentation approaches are covered. In Chapter 3, our modified and improved approach to solve the mitosis detection task is illustrated. In chapter 4, we introduce the MITOSIS dataset used in our experiments and the way we process it. The evaluation metrics and the implementation details are introduced. Eventually, the experimental results are presented and analyzed. In Chapter 5, we introduce our real-time computational tool and the "correction and relearning" pipeline in detail. At last, the conclusions of the research and future work are discussed in Chapter 6.

Chapter 2

Literature Review

In this chapter, firstly, we will describe the background knowledge of Convolutional Neural Networks (CNN) and several popular CNNs. Secondly, we will review the difficulties in detecting mitotic cells, and some previous works on the mitotic cells detection tasks. Thirdly, we will describe the You Only Look Once (YOLO) series object-detection frameworks. Finally, we provide an overview of different copy-paste data augmentation methods.

2.1 Convolutional Neural Network

Convolutional Neural Network (CNN) is a type of deep neural networks. It is designed to automatically and adaptively generate spatial hierarchies of feature maps. It consists of a sequence of layers. Each layer of a CNN transforms to another CNN through a differentiable function. There are three main layers in a CNN including Convolutional layers, Pooling layers and Fully-Connected layers. These layers will be stacked to build a CNN. The CNN structure is shown in figure 2.1.

Convolution layer

Convolutional layers aim to extract features from an input image by calculating the convolution of a filter/kernel and a region of the input image. In general, the first few layers extract shallow

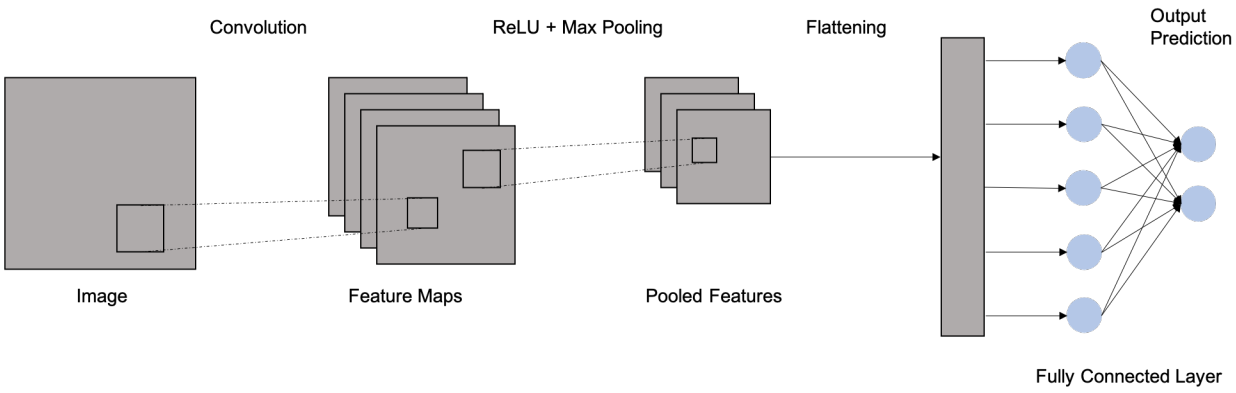


Figure 2.1: Convolutional Neural Network

level features, for instance, edges and shapes of objects. The subsequent layers are responsible to extract more high level abstract features. These shallow and high level features can be utilized to classify images or identify objects in an image. In addition, researchers sometimes shift the filters by different number of pixels over the input matrix to obtain different information. Stride is used to describe the number of shifted pixels. Therefore, the filters will move 1 pixel if the stride equals to 1. The following figure 2.2 illustrates the convolutional procedure. The convolution operation preserves the relationship between pixels.

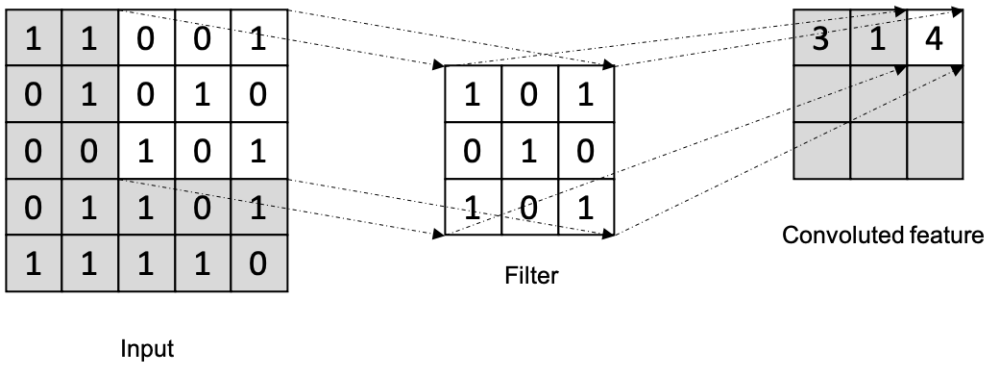


Figure 2.2: The convolution operation to generate a feature map.

Pooling layer

Spatial pooling aims to reduce the size of each feature map but retain the spatial information. It reduces the number of parameters and boosts the computational efficiency. Some common pooling operations are max pooling, average pooling, and sum pooling. These pooling oper-

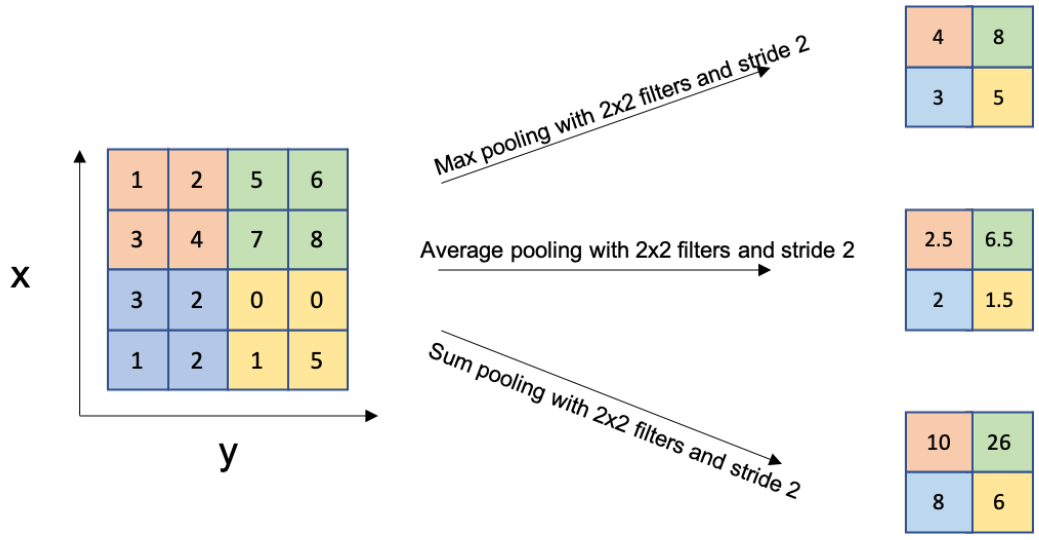


Figure 2.3: Pooling Layer

ations produce more robust feature maps. The following figure 2.3 demonstrates these three common pooling operations with a 2×2 filter and stride 2. Max pooling takes the largest element from the calculated feature map. As shown in figure 2.3, the value is 4 in the upper left corner, the value is 8 in upper right corner, and so on. Average pooling and sum pooling take similar operations which takes average value in an area, and sums the values in an area respectively.

Fully Connected Layer

After processing the convolution operations, the final feature map will be flattened into vector and feed it into a fully connected layer like a neural network. The entire progress is illustrated as shown in the figure 2.1. After flattening the pooled feature map, the flattened vector will be feed into a artificial layer neural network. Eventually, the final layer will predict a probability for each output neuron with different activation functions.

The CNN has been widely used in different computer vision domains such as the image classification and object detection task. Some popular CNNs are VGG-16 [43], ResNet-50 [17], and DarkNet[37].

VGG-16 achieves the second place in ImageNet competition in 2014. The default input size

of VGG-16 is 224 × 224. It has 16 layers including 13 convolutional layers and 3 FC layers. The filter size of VGG-16 is 3 × 3 with stride equals to 1. The pooling kernel is 2 × 2 with stride equals to 2. VGG-16 significantly reduces the size of the kernel size and increases the number of convolutional layers. However, VGG-16 consumes more computational resources due to the 3 FC layers which contains exceeded parameters.

With the development of CNN, there are two problems arising while deepening of the network structure. The first problem is vanishing gradient. It can be overcame by using the normalized initialization. The second problem is degradation. It describes the situation that the accuracy decreases while deepening the network structure. Some information will loss due to the convolutional operations and pooling operations in the convolutional layer. Therefore, ResNet-50 has been proposed to address this problem by using a deep residual learning block. ResNet-50 wins the champion in the 2015 ImageNet competition. The block is shown in the figure 2.4. Authors directly add the value of shallow layer with the value after the convolutional operations which is defined as a "short-cut". Specifically, ResNet is composed of different bottleneck as shown in the figure 2.5. In the bottleneck structure, the authors firstly utilize a 1 × 1 convolution filters to proceed the convolutional operation which aims to reduce the dimension of channels. Afterwards, a 3 × 3 convolution filter is applied to extract features. Finally, another 1 × 1 convolution filter is applied to recover the original dimension, and the short-cut is applied to pass the information from shallow layers. In general, ResNet contains 5 groups of convolutions. The input size of ResNet is 224 × 224, the output size of ResNet is 7 × 7. Compared with the VGG-16 and other traditional CNNs, ResNet can not only reduce the computational resources caused by the exceeded parameters in FC layers in VGG-16, but also performs better with more layers.

DarkNet-53 utilizes the residual block as ResNet to build a deep CNN. It integrates the L2 regularization into the convolutional operations. After each of the convolutional operations,

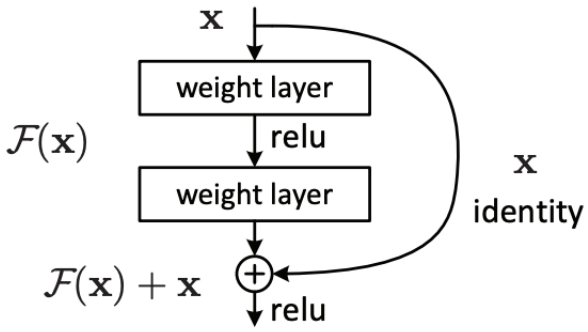


Figure 2.4: ResNet Residual Learning block [17]

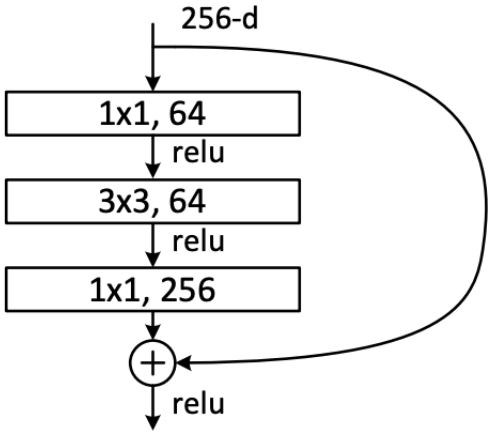


Figure 2.5: Bottleneck of ResNet[17]

DarkNet-53 uses LeakyReLU as the activation function instead of ReLU in traditional CNNs.

$$ReLU(x) = max(0, x) \tag{2.1}$$

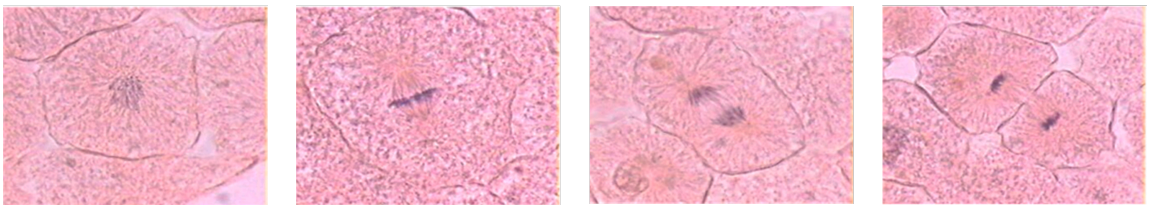
$$LeakyReLU = \begin{cases} x & \text{if } x \geq 0 \\ \frac{x}{\alpha} & \text{otherwise} \end{cases} \tag{2.2}$$

These CNNs are widely used in image classification tasks and also used as backbone networks to extract features in object detection tasks. However, in the object detection tasks, the FC layers are removed. Feature maps from the final layer or from different layers will be utilized directly to do further processing to generate the detection results.

2.2 Mitotic Cells detection

Researchers have proposed many mitotic cells detection methods, but existing methods are still unable to achieve satisfactory results in clinical practice. It is difficult to detect mitosis in pathological images of breast cancer for the following reasons.

- **Different stages of mitosis**: Mitosis can be categorized into four stages: prophase, metaphase, anaphase and telophase. The morphological characteristics in each stage



(a) Prophase (b) Metaphase (c) Anaphase (d) Telophase

Figure 2.6: Four stages of mitosis. (a) Prophase. (b) Metaphase. (c) Anaphase. (d) Telophase [33]

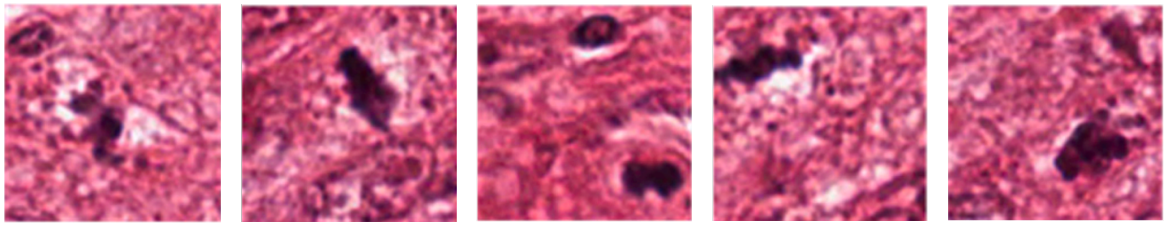


Figure 2.7: Examples of mitotic cells [33]

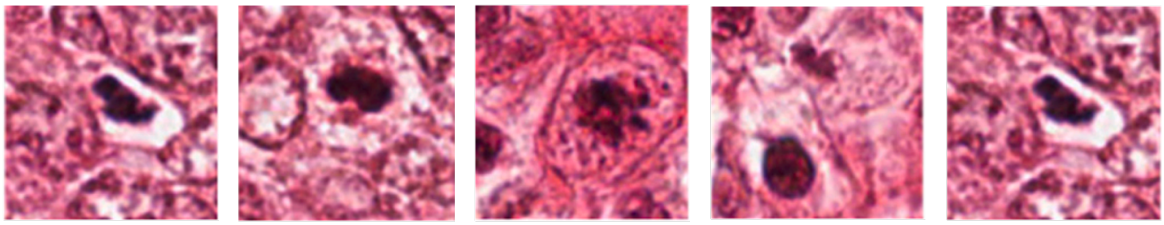


Figure 2.8: Examples of the non-mitotic cells [33]

are significantly varied. Examples are shown in the figure 2.6. In telophase, there are two nuclei. However, they need to be counted as a single cell since they have not yet divided into two separated cells.

- **Similarity between mitotic and non-mitotic cells**: The morphology features of mitotic cells is diverse, the morphology of some non-mitotic cells are similar to mitotic cells. As shown in the figure 2.7 and figure 2.8. For example, some apoptotic cells produce great interference, which also complicates the accurate detection of mitotic cells [33].

- **Low density of mitotic cells**: The density of mitotic cells is extremely low in a single image. In an pathological image, the number of mitotic cells is far less than non-mitotic cells. Examples are shown in the figure 2.9 with annotations. It is demonstrated in the

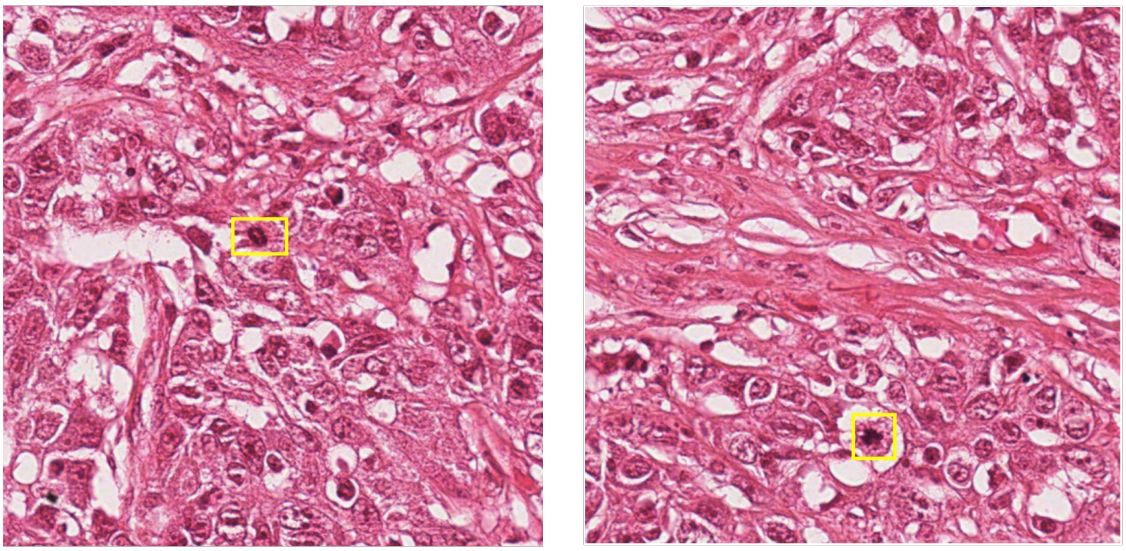


Figure 2.9: Unbalanced positive and negative samples

example figures that the number of mitotic cells are extremely rare.

These factors make significantly impacts on extracting features and detecting accurately in mitotic cells detection task. These factors must be dealt with some additional methods to ensure the accuracy and robustness of mitotic cells detection. Many works focus on providing automatic solutions to mitotic cell detection. These works can be mainly divided into two categories according to the features of region-of-interests (ROIs), namely handcrafted-features based and deep-fetures based. Details of each of the aforementioned categories are introduced as follows.

2.2.1 Handcrafted-Features Based Motitic Cell Detection

A handcrafted-features based mitotic cell detection method is generally involved into three steps. First, plenty of multi-scale sliding windows are utilized to scan the entire input image and generate region proposals. However, this approach is computationally expensive and generates many redundant windows. After these sliding windows are generated, the second step is to extract some representative features used for mitotic cells detection through SIFT [29], HOG [9] and Haar-like [26]. Texture, shape, and corner are the major representative features which

are designed based on the domain knowledge of pathologists. These features can provide semantic and robust representations based on these region proposals. However, due to the diversity of target shapes, colors, lighting conditions, and backgrounds, it is more difficult to design relevant robust features. Eventually, these representative features will be feed into a robust classifier to distinguish a target object from all the other categories.

In the mitosis detection challenge of ICPR MITOSIS 2012, the approach [19] proposed by Irshad has ranked second, and the technique [47] presented by Tashk et al. has ranked third. The former approach starts by segmenting target objects. Further, the author utilizes a decision-tree classifier [30] to generate classifications based on the extraction of statistical and morphological features. The latter method adopt local binary pattern (LBP) and SVM [30] as the classification algorithm. However, LBP features are invariant to grayscale changes. Sommer et al. present a pipeline [44] to distinguish between mitotic and non-mitotic cells. The authors employ two open-source biomedical softwares named *ilastik* [45] to segment objects, and *CellCognition* [18] to provide the binary classification results. In their proposed technique, the combination features of shape, intensity and texture are feed into a SVM classifer to produce the final classification result. However, compared to [19, 47], this technique makes a low detection performance. Paul et al. propose a technique [34] which achieves a significant detection performance based on a regenerative random forest tree classifier. They focus on the nucleus of the mitotic cell instead of the whole cell. However, this proposed technique can not be accepted into a real-time application due to the high demanding computational resources.

However, the manually designed features can not be representative for all mitotic cells. Recently developed deep-features based techniques are more powerful than handcrafted-features based methods. Since deep-features based methods can extract plenty of features during training instead of manually designing the features.

2.2.2 Deep-Features Based Mitotic Cells Detection

Ciresan et al. utilized the sliding windows approach to extract the deep features [6]. This method achieves the highest F-Score on the ICPR MITOSIS 2012 dataset. However, this technique is not suitable for real-time clinical applications due to its computationally expensive on generating sliding windows. Malon et al. present a method [32] to combine deep features generated from a CNN and the handcrafted features together. This method holds a high computational complexity, and the performance is not satisfactory. Wang et al. propose a cascade technique [48]. They use conventional image processing methods to select regions of interest (RoI), and train two independent classifiers for handcrafted features and deep features respectively. During the testing phase, if two classifiers generate two different results for an image, an additional classifier is utilized to make a final decision. This additional classifier is trained with both features. This cascade technique requires less computational resources. However, the proposals are selected by the conventional cell segmentation method, which is not reliable and effective. Afterwards, Chen et al. propose a cascaded convolutional neural network [4] to detect mitosis. This method is composed of two components. First, a fully convolutional network (FCN) is developed to generate region proposals. Second, an additional CNN is adopted as a classifier to classify mitotic cells from these proposals. However, these two networks are trained separately, and it is still time-consuming to construct the handcrafted-features and combine them with deep-features together.

Instead of selecting region proposals based on handcrafted features and training network separately, many researchers have utilized Region-based convolutional neural network (R-CNN) methods in the mitotic cells detection task to implement an end-to-end model and obtain better results. These R-CNN based methods demonstrate impressive performance in many computer vision applications. The first R-CNN is proposed by Ross B. Girshick in 2014 [13]. It adopts the selective search algorithm to divide an input image into different proposal regions. It utilizes a pre-trained CNN model based on a large dataset (such as ImageNet ILSVC 2012) to extract the features of these proposal regions. Eventually, the final feature map is sent to

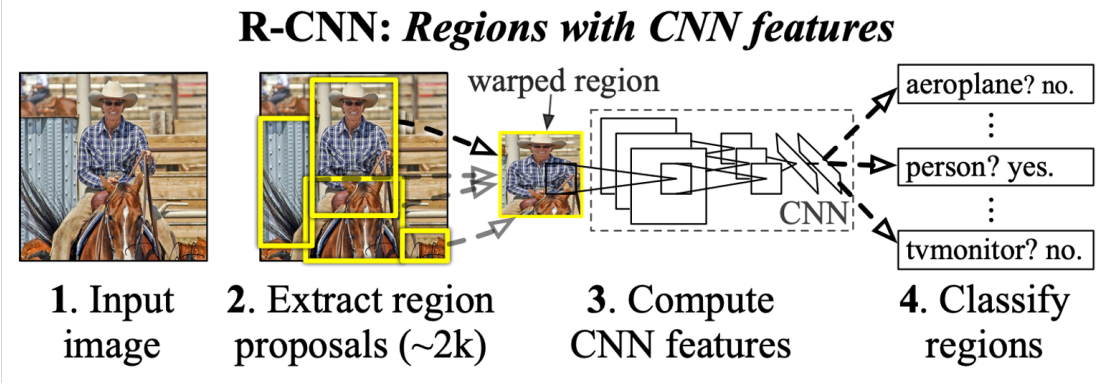


Figure 2.10: R-CNN [13]

a SVM classifier to generate class labels, and a regression network is trained simultaneously to refine the bounding box of proposal regions which aims to minimize the distance between ground truth and predicted bounding boxes. R-CNN model achieves 31.4% mean Average Precision on the ILSVRC 2013 dataset, which is significantly ahead of the second OverFeat [42] method by 24.3%. The whole procedure is illustrated in figure 2.10.

The selective search procedure has become the bottleneck of R-CNN, therefore, Ren et al. proposed the Faster R-CNN [39] which significantly improves the detection accuracy and speed based on R-CNN. They generate region proposals by a neural network rather than utilizing the selective search procedure. The procedure of Faster R-CNN starts with extracting features, the final layer feature map is shared to be used in the Region Proposal Network (RPN) and RoI pooling layer. The RPN substitutes the selective search method in R-CNN and Fast R-CNN, and is proposed to generate region proposals. It generates 9 anchor boxes for each pixel in the features map. RPN adopts softmax function to determine whether an anchor box belongs to a foreground or a background, and obtain the accurate region proposals by bounding box regression. Further, the proposal feature maps will be extracted by RoI pooling layer from both region proposals and the feature map. RoI pooling layer generates fixed-length feature vectors and send them to the fully connected layer for final classification. Eventually, the categories of the region proposals are calculated, and bounding box regression is adopted again to refine the position of the region proposals. The general procedure is shown in the figure 2.11. Faster

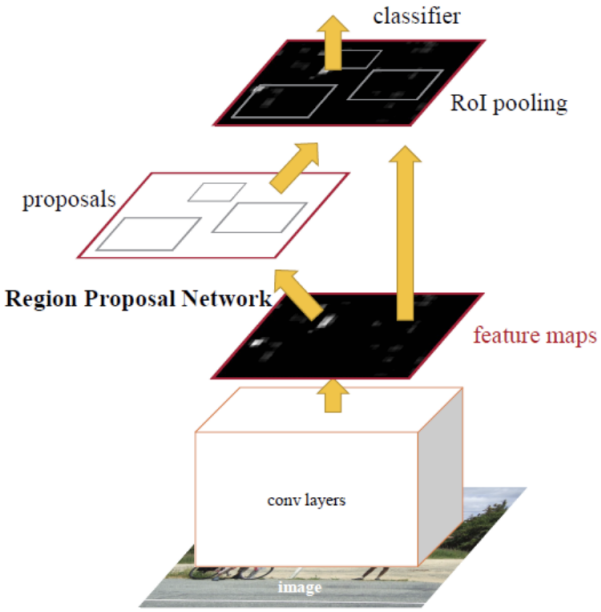


Figure 2.11: Faster R-CNN [39]

R-CNN preserves the Non-maximum Suppression (NMS) approach to get rid of the redundant bounding boxes, which makes sure the object detection algorithm only detects each object once.

Li et al. [23] firstly utilize the Faster R-CNN [39] with visual geometry group (VGG)-16 [41] as a feature-extraction network to detect mitotic cells. They also utilize an additional residual network (ResNet)-50 [17] as the further classification network to refine the detection results. This technique achieves the highest F1-score on the ICPR MITOSIS 2012 challenge with less inference time. Afterwards, Cai et al. develop an improved method by replacing the backbone network with ResNet 101 to extract features based on the studies of Li et al [2]. Dodballapur et al. [10] utilize mask R-CNN [15] with ResNet-50 as the backbone network to extract features in pixel level. In order to reduce the false positives generated from the mask R-CNN, a Xception network [5] has been adopted. This method achieves a high detection performance. However, it can not be used in the real-time application due to the expensive in the training procedure. Mahmood et al. proposed a four-stages mitotic-cell detection framework [31]. In the first stage, Faster R-CNN with the ResNet-50 as the backbone network is

exploited to detect mitotic cells. There are many false positives generated by Faster R-CNN. They perform post-processing on the basis of statistical, texture, shape, and color features to reduce the number of false positives which are caused by the differences between mitotic and non-mitotic objects. In the third stage, they perform a score-level fusion of Resnet-50 and a dense convolutional network (DenseNet)-201 to further reduce the false positives. Eventually, the final classification of the mitotic and non-mitotic cells is performed in stage 4. Their result achieve the excellent detection performance. However, owing to its use of expensive GPUs and intensive training, it is not suitable for use in practical applications. All of these strategies developed based on R-CNN can be improved by integrating various feature pyramid networks and using powerful backbone networks.

2.3 You Only Look Once (YOLO)

Faster R-CNN is a two-stage end-to-end detection method. It first generates region proposals via the Region Proposal Network (RPN). Afterwards, it generates the classification and regression results based on these proposals. However, generating region proposals on a histopathological slide is extremely time-consuming due to its competitive size. Compared with two-stage algorithms, one-stage methods directly predict class probabilities and bounding box offsets from input images without generating region proposals. It is an end-to-end detection system that encapsulate all stages in a single neural network. One of the representative one-stage object detection methods is You Only Look Once (YOLO) [36]. So far, YOLO has released different versions [36, 37, 38, 1]. YOLOv1 laid the foundation for the entire series. Latter versions are developed to improve the performance of YOLOv1.

YOLOv1 divides an image into different $\mathbf{S} \times \mathbf{S}$ grids. If the center of an object is located in a grid, this grid cell is responsible for detecting the object. Furthermore, each of those grid cells predicts **B** number of bounding boxes and the corresponding confidence scores of these boxes. The confidence scores each contain the probability that each box contains an object as

well as the positional accuracy of the predicted box. The macro pipeline is shown in figure 2.12.

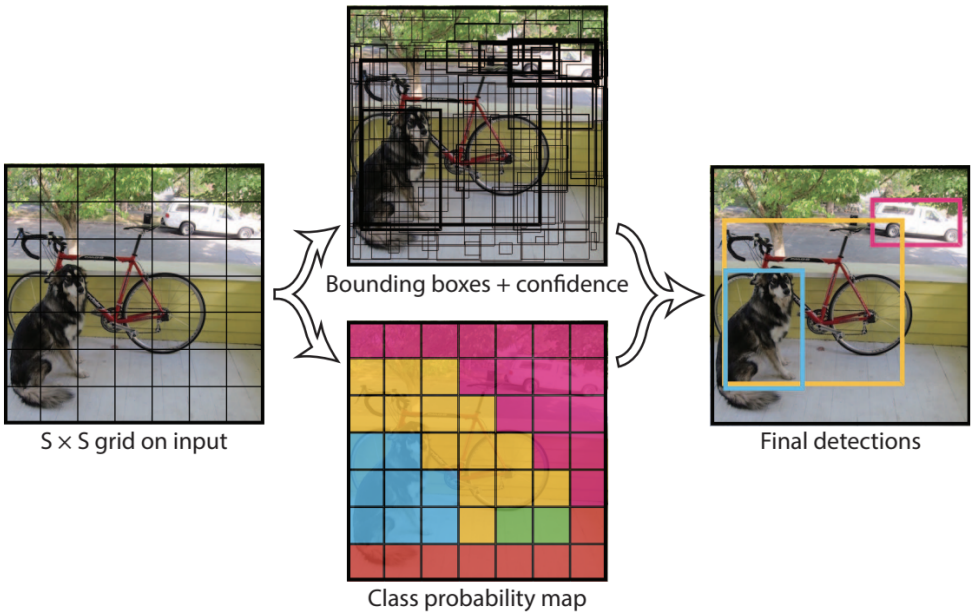


Figure 2.12: An overview of YOLO detection method [36]

YOLOv1 is similar to R-CNN in the extraction of features. Both of them find bounding boxes in the grid cells and extract features via a Convolutional Neural Network. However, in YOLOv1, each grid is responsible for predicting one class. These spatial restrictions are added to the grid cells in YOLOv1, which helps to prevent duplicate detection of the same object. However, YOLOv1 infers faster but less accurately than Faster RCNN. YOLOv2 has some improvements by utilizing more data and adopting a new network to extract features [37]. However, the accuracy is still worse than Faster RCNN. YOLOv3 [38] improves the detection accuracy dramatically by using different tricks, such as utilizing a deeper backbone network named Darknet53 to extract significant features and utilizing the Feature Pyramid Network (FPN) to generate multi-scale predictions. FPN algorithm substitutes the feature extractor in YOLOv2 by generating multiple feature map layers.

FPN involves a bottom-up and a top-down pathway, as shown in the figure 2.13. The bottom-up pathway follows the usual convolutional network to extract features. The spatial

image resolution decreases while more high-level features are constructed and the semantic value for each layer increases. The top-down pathway is executed after the final feature map from bottom-up pathway is accessed. The upsampled map is then merged with the corresponding bottom-up map by element-wise addition, and generate prediction results on each feature layers. FPN significantly increases the COCO-style Average Precision (AP) by 2.3% and PASCAL-style AP by 3.8%, over a strong single-scale baseline of Faster R-CNN on ResNets.

.

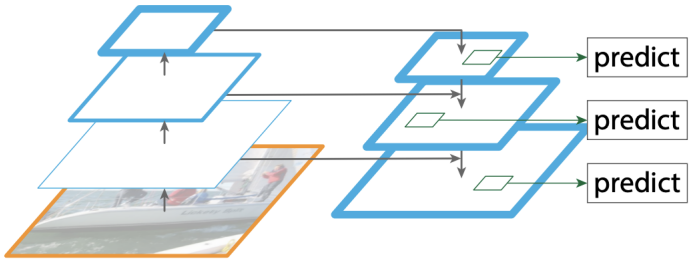


Figure 2.13: Feature Pyramid Network (FPN) [27]

These two major modifications in YOLOv3 significantly improve the detection performance than YOLOv2. However, it traded off against speed which reduced from 45 to 30 frames per second (FPS) due to the depth of backbone network. Recently, YOLOv4 have been published. It integrates multiple advanced tricks to achieve a competitive result. It utilizes an improved backbone network to extract features. In addition, instead of utilizing FPN [27] in YOLOv3, YOLOv4 use Spatial Pyramid Pooling [16] and PANet [28] to extract features. PANet is proposed to utilize feature maps efficiently. It adds a bottom-up path to augment the top-down path in FPN, and utilizes the adaptive feature pooling to capture information from all levels.

Compared with Faster R-CNN and YOLOv3, YOLOv4 performs better and faster in the popular object detection datasets, including MS COCO and PASCAL VOC dataset. Though there have some methods to detection mitotic cells, we have observed that no any one-stage detection network, like YOLO, has been applied to this task. Different from previous methods, we modify and propose a lightweight detection model based on YOLOv4 to solving mitotic

cells detection problem.

2.4 Data Augmentation

In the real world scenario, the number of annotated training image for object detection task is in short supply due to the expensive cost of manually annotating objects in every image. Therefore, researchers use different data augmentation strategies to pre-process training data. Data augmentation strategies are essential to effectively enlarge the dataset for free and reduce overfitting without modifying the ground-truth labels. Common data augmentation strategies such as random crop [20, 21, 43, 46], color jittering [46] and Auto/RandAugment [7, 8] are integrated into image classification tasks to achieve better performance on the ImageNet benchmark [40]. These augmentations are mainly used for encoding invariances to data transformations, which are more suitable for image classification tasks[40].

2.4.1 Mixing Image Augmentations

In addition to the aforementioned data augmentations, there is another branch of data augmentation strategies that mixes the information contained in different images and make appropriate adjustments to ground-truth labels. One classic example is the MixUp data augmentation [50] method. It creates a new image by blending a pair of images from training set under different weight. CutMix [49] is one variations of mixup. Instead of mixing pixels of two images, CutMix pastes rectangular crops of an image into another image. Examples of MixUp and CutMix is shown in figure 2.14.

Both MixUp and CutMix are utilized in object detection task [51]. The Mosaic data augmentation method employed in YOLOv4 [1] is related to CutMix. Mosaic data augmentation combines multiple individual training images with different ratios into a new compound image along with their ground-truth labels. Examples are shown in the figure 2.15. Mosaic data augmentation strategy allows the model to identify objects at a smaller scale than normal.

Figure 2.14: Left: MixUp. Right: CutMix

Figure 2.15: Mosaic data augmentations[1]

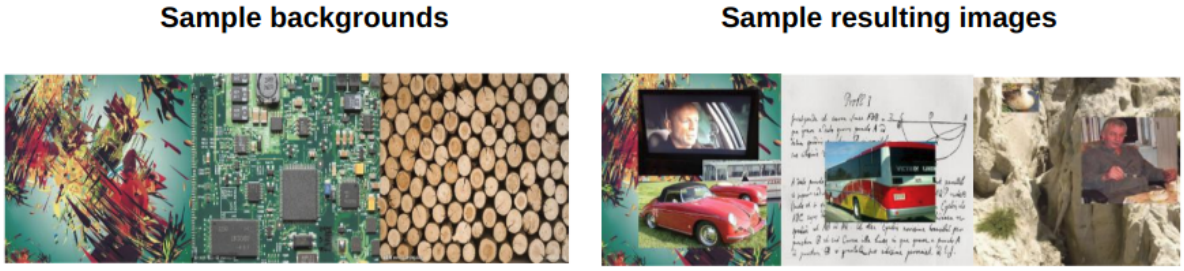


Figure 2.16: Examples of cut-and-paste [35]

However, it is not suitable in mitotic cells detection task. It will be difficult to detect mitotic cells if the scale is relatively small. Although MixUp, CutMix and Mosaic are good at creating new training data by combining multiple images, these methods are still not object-aware and have not been designed specifically for the task of instance segmentation.

2.4.2 Copy-and-Paste Augmentation

Rao et al. propose a straightforward approach [35] on ICVIP in 2017. They cut objects from VOC2007 and VOC2012 object detection datasets and paste these objects onto images with different backgrounds. Examples are shown in the following figure 2.16. They train with these images, and it turns out that the results of standard one stage object detection network like YOLO have been significantly improved with less wrong labels and more accurate bounding boxes.

A similar but slightly less naive approach to cut and paste was introduced by researchers

from the Carnegie Mellon University in ICCV 2017 named Cut-Paste-and-Learn [11]. Authors extract object instances and collect different background scenes. They blend and paste these instances on different scenes and train with the augmented images. Eventually, they test on real images. A graphical overview of their approach is shown in figure 2.17.

In addition, Google research team [12] propose a simply copy-and-paste technique to generate new training data. Firstly, two images are randomly selected, and they apply random scale jittering and random horizontal flipping on both images. Secondly, the authors randomly copy a subset of objects from one image and paste them onto the other image without modeling the surrounding context. However, this action will create some fully or partially occluded objects. Eventually, authors remove fully objects and update the masks and bounding boxes of partially occluded objects. A sample is shown in figure 2.18. The experimental results are claimed that pasting objects randomly can provide solid improvements on top of strong baselines.

Figure 2.17: Examples of cut-paste-and-learn [11]

Figure 2.18: Examples of simple Copy and Paste [12]

Chapter 3

Methodology

In this chapter, we will explain our approach towards solving the difficulties in mitosis detection task in detail. Firstly, the structure of our modified lightweight YOLOv4 is introduced. Secondly, different loss functions are introduced to take responsibility to detect small, medium and large objects. A different method to suppress the redundant bounding boxes is presented. In addition, we also train our modified YOLOv4 model with similar data augmentation method mentioned in [12]. In [12], authors randomly crop objects in pixel level from two images and paste these cropped objects at another image. However, we have limited images for training. We crop mitotic cells at instance level instead of pixel level, and we paste these cropped cells back at the same image by randomly flipping and rotating them. This proposed method can increase the number of positive samples and alleviate the poor detection performance. Eventually, our model is applied to detect the mitosis and obtain estimated bounding boxes annotations on the MITOSIS and WHPG dataset.

3.1 Modified YOLOv4

Since the resolution of a histopathological slide is extremely large, it is not convenient to train the model with these slides. Therefore, we crop small patches from original slides using the same method mentioned in [23] and [31]. The architecture of our modified YOLOv4 detection

network for the mitotic cells detection task is shown in Figure 3.1.

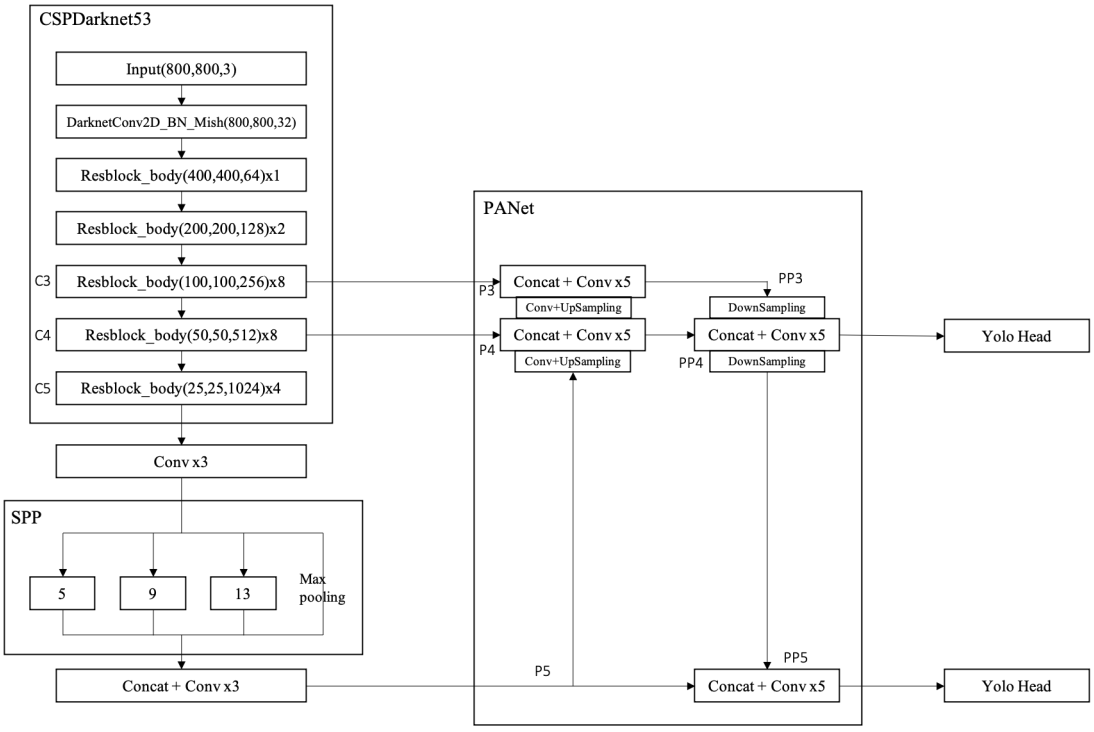


Figure 3.1: Architecture of our modified network

3.1.1 Detection Model

YOLOv4 consists two major components: feature extractor and multi-scale detector. The small patch is first fed into feature extractor to extract feature embedding. YOLOv4 uses CSPDarknet53 as the backbone network to extract features. CSPDarknet53 is a convolutional neural network which is developed by the combination of Darknet53 and CSPNet. Darknet53 is the backbone network utilized in YOLOv3, as shown in the figure 3.2. CSPNet optimizes the problem of repetition of gradient information in backbone network for feature extraction by spliting the feature map of the base layer into two parts and merging them through a cross-stage hierarchy. As shown in the figure 3.3. It ensures the speed and accuracy of inference and decreases the model size by reducing the number of parameters.

Additionally, the Spatial Pyramid Pooling (SPP) block is adopted in YOLOv4. It replaces the last pooling layer of CSPDarknet53 with a spatial pyramid pooling layer. In the last layer

of CSPDarknet53, YOLOv4 utilizes four kernels with the size of 1×1, 5×5, 9×9 and 13×13 to proceed the maxpooling. The spatial dimension is preserved. The features maps from different kernel sizes are then concatenated together as the output. This output is utilized further in the multi-scale detector. The purpose of SPP is to increase the receptive field and obtain the most significant contextual features.

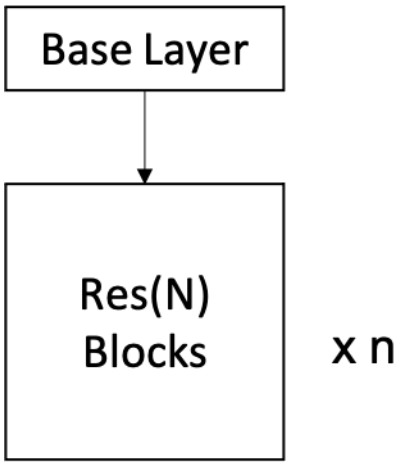


Figure 3.2: Darknet53 in YOLOv3

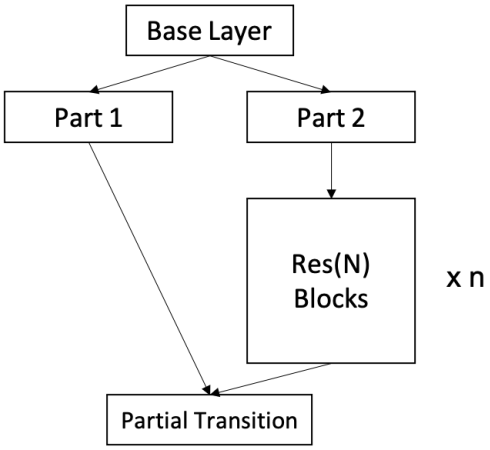


Figure 3.3: CSPDarknet53 in YOLOv4

Except for the feature extractor, YOLOv4 modified the Path Aggregration Network (PANet) to extract features iteratively instead of utilizing FPN in YOLOv3. YOLOv3 only adopt the feature maps P3, P4 and P5 to generate the detection results. However, as an image passes through the CSPDarknet53, the spatial resolution of the feature maps decreases due to various strided convolutions and pooling layers. The decrease in spatial resolution leads to a loss of spatial information in the shallow layers. Therefore, besides the top down path used in FPN, PANet suggests a bottom-up path augmentation which iteratively combines shallow features with high level features. The original PANet[28] is slightly modified in YOLOv4. Instead of adding the neighbouring layers, YOLOv4 applies a concatenation operation on these neighbouring layers to improve the performance of predictions.

Originally, YOLOv4 is designed and tested on the MS COCO and PASCAL VOC dataset. The sizes of objects in both dataset vary from 30x30 to 300x300 pixels. Therefore, authors generate three different predictions on three different heads with different scales in order to

cover the sizes of objects as much as possible. The layer used to detect and get the final output of Darknet is called the YOLO head. However, a mitotic cell is relatively small, and the sizes of mitotic cells do not change significantly in the ICPR MITOSIS datasets compared to the sizes of objects in MS COCO and PASCAL VOC datasets. Therefore, three prediction heads are redundant in our case. In this work, we only utilize two prediction heads for the mitotic cells detection task.

We generate three effective feature maps C3, C4 and C5 with different shapes. The down-sampling (bottom-up) pathway is the feed-forward computation of the CSPDarknet35. Specifically, the spatial resolution is down-sampled by a factor of 2 using the nearest neighbor. The feature maps from down-sampling pathway undergoes 1×1 convolutions to reduce the channel dimensions. After all the feature maps are calculated, three effective feature maps C3, C4 and C5 from the bottom-up pathway and the top-down pathway are merged through element-wise addition. These transformed feature maps P3, P4, P5 are generated from the merged feature maps which undergo 3×3 convolution to reduce the aliasing effect of up-sampling. The features P3, P4, P5 are merged through element-wise addition as well. Eventually, only PP4 and PP5 are utilized in the prediction heads for predicting bounding boxes and correlated classes.

3.1.2 Loss Function

YOLOv4 generates the bounding boxes, confidence scores and the class probabilities in each prediction head. The loss function of YOLOv4 involves confidence loss, classification loss, and regression loss [1]. These loss functions are calculated on different prediction heads. There are two different YOLO heads in our model, therefore, the final loss is calculated by summing the losses calculated by these two heads and then carry out back propagation.

The losses are calculated based on the predicted bounding boxes. Therefore, we utilize the concept of Intersection over Union (IoU) to evaluate the performance of bounding boxes predictions. It computes intersection over the union of two bounding boxes. As shown in the figure 3.4. If an IoU value equals to 1, it implies that these two bounding boxes are perfectly

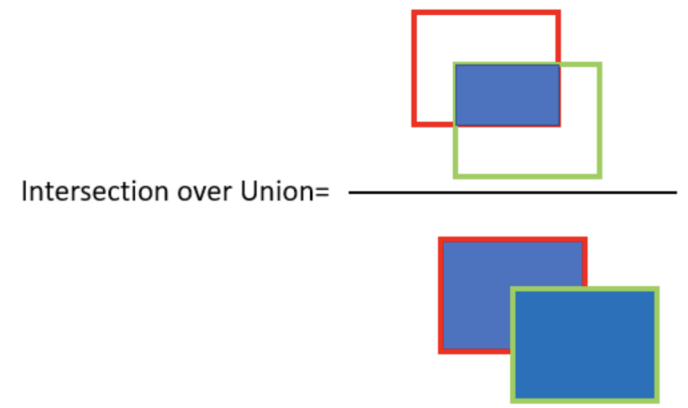


Figure 3.4: Intersection of Union

overlap. In YOLOv4, for each object, there are many predicted bounding boxes and one ground truth bounding box. For each predicted bounding box, an IoU is calculated with correlated ground truth bounding box. The predicted bounding box with the largest IoU is selected as a positive example for each object. In YOLOv4, there is only one positive bounding box for each object. Therefore, in order to balance the dramatic difference between the number of positive and negative samples, authors employed the ignored bounding boxes. Beside of the positive bounding boxes for each object, other predicted bounding boxes which IoU is greater than a threshold will be considered as ignored bounding boxes. These ignored bounding boxes do not contribute to any losses. Besides the positive bounding boxes, the IoU of a predicted bounding box below a threshold value is considered as a negative bounding box. The negative bounding boxes only contribute to the confidence loss.

Complete Intersection over Union (CIoU) [52] loss function is utilized as the regression loss in each yolo head for only positive bounding boxes. However, according to the [52], the performance of CIoU on small objects has decreased. The aspect ratio does not contribute much to the detection of small objects. This could be one possible reason why CIoU performs worse in detecting medium and small objects. The center point is more important than the aspect ratio for small objects. For the mitotic cells detection tasks, the sizes of most mitotic cells are small. Therefore, we use the Distance IoU (DIoU) loss as the regression loss function for medium and small objects, CIoU loss as the regression loss function for the relative large

objects. The formulas of CIoU and DIoU loss are expressed as follows.

$$L_{DIoU} = 1 - \frac{b \cap b^{gt}}{b \cup b^{gt}} + \frac{p^2(b, b^{gt})}{c^2} \tag{3.1}$$

$$L_{CIoU} = 1 - \frac{b \cap b^{gt}}{b \cup b^{gt}} + \frac{p^2(b, b^{gt})}{c^2} + \alpha v \tag{3.2}$$

where b^{gt} is the ground truth bounding boxes, b is the predicted bounding boxes. p denotes to the Euclidean distance, c denotes the diagonal length of the smallest enclosing box covering the two boxes. α is a positive trade-off parameter, and v measures the consistency of aspect ratio.

$$v = \frac{4}{\pi^2}(arctan\frac{w^{gt}}{h^{gt}} - arctan\frac{w}{h})^2 \tag{3.3}$$

We use the same loss functions for classification and confidence loss as the original YOLOv4. YOLOv4 adopts the cross entropy loss function to calculate confidence loss and classification loss. Cross entropy loss function is used to evaluate the difference between the ground truth and the prediction. The ground truth and the prediction is closer while the value of cross entropy is smaller. For the positive example, we calculate the confidence loss by using the cross-entropy loss. This contains two parts, if the bounding box contains an object or not. The following formula is used to calculate the confidence loss.

$$\begin{aligned} L_{conf} = &\sum_{i=0}^{S^2} \sum_{j=0}^{B} I_{ij}^{obj}[\hat{C}_i log(C_i) + (1 - \hat{C}_i) log(1 - C_i)] - \\ &\lambda_{noobj} \sum_{i=0}^{S^2} \sum_{j=0}^{B} I_{ij}^{noobj}[\hat{C}_i log(C_i) + (1 - \hat{C}_i) log(1 - C_i)] \end{aligned} \tag{3.4}$$

where S is the size of grid, B is the number of bounding boxes in each grid, if the bounding box at position (i,j) contains an object, I_{ij}^{obj} equals to 1, otherwise, equals to 0. Similarly, if the bounding box at position (i,j) does not contain an object, I_{ij}^{noobj} equals to 1, otherwise, equals to 0. If this bounding box is positive, then $\hat{C}_i$ equals to 1, otherwise, it equals to 0. For

classification loss, the following formula is used.

$$L_{cls} = \lambda_{class} \sum_{i=0}^{S^2} \sum_{j=0}^{B} I_{ij}^{obj} \sum_{c \in classes} [\hat{p}_i(c) log(p_i(c)) + (1 - \hat{p}_i(c)) log(1 - p_i(c))] \tag{3.5}$$

where p_i denotes as the predicted class probability, if the predicted class belong to the correct class, then $\hat{p}_i(c)$ equals to 1, otherwise, equals to 0. Eventually, the total loss equals to L_{conf} + L_{cls} + L_{CIoU} + L_{DIoU} but with different weights of each loss.

3.1.3 DIoU Non-Maximum Suppression

During the testing phase, we adopt the DIoU Non-Maximum Suppression method to post-process the redundant bounding boxes. Non-Maximum Suppression (NMS) is a commonly used technique in the post-processing of one-stage and two-stage object detection algorithms to filter out the redundant predicted bounding boxes. The following is the process of selecting the best bounding box through NMS. The first step is selecting the bounding box with the highest confidence score. The second step is comparing the intersection over union of this boxes with other bounding boxes. Third, we need to remove the bounding boxes based on the value of intersection over union which is greater than a threshold. Afterwards, we choose the next bounding boxes with the second highest confidence score, and repeat the step 2 to 4 until all the bounding boxes are filtered.

However, the classic NMS method only considers the IoU and does not consider the context information. DIoU can be integrated into non-maximum suppression (NMS) as the criterion. Therefore, DIoU-NMS [52] not only considers the IoU, but also considers the distance between the center points of the two bounding boxes. For example, if the IoU between two boxes is relatively large, and the distance between the two boxes is relatively large as well, these two boxes need to be considered as two separate objects.

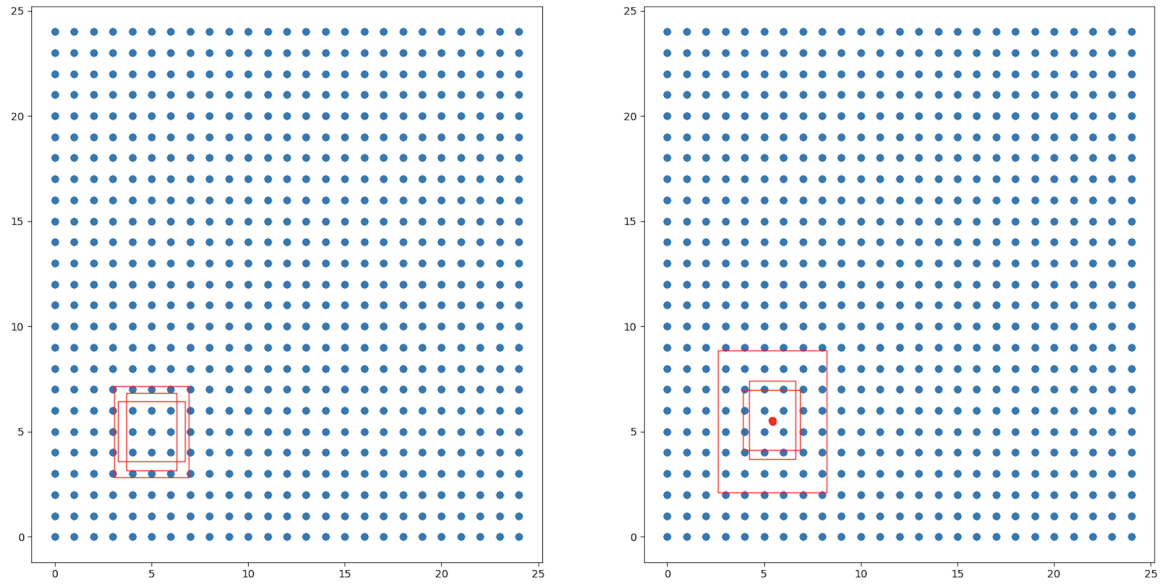


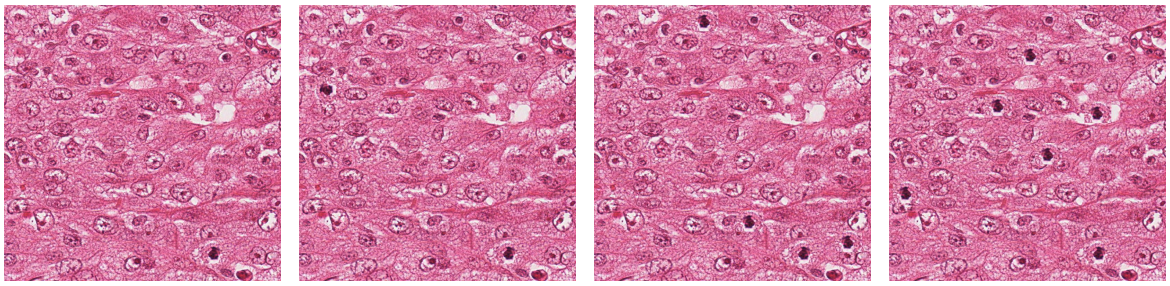
Figure 3.5: An example of Anchor boxes in 25x25 scale feature maps. Left: Anchor boxes with constant sizes and scales. Right: Anchor boxes with adjusted sizes and scales based on the ICPR MITOSIS dataset.

3.1.4 Anchor Boxes Adjustment

In order to allow one grid to detect multiple objects, we pre-defined some boxes with different scales and aspect ratios at each position. These boxes are called anchors. The aspects of the anchors strongly depends on the dataset being used, and the detection performance can be various by different sizes and scales of anchor boxes. In the following Figure 3.5, a set of anchors defined in 25×25 feature maps from our dataset is illustrated. The sizes and scales of anchor boxes are calculated by K-Means in YOLOv4 based on the ground truth bounding boxes in the training set.

3.2 Copy-and-Paste Augmentation

In our work, inspired by [12], we utilize an similar approach to copy and paste cells. Instead of cropping objects in pixel level, we crop the cells in instance level. When cropping cells from an pathology image, we only consider the cells without occluding with other cells to prevent the image from being too unreal. When pasting these cells to the same pathology image, we



(a) Original Images (b) One time (c) Three times (d) Five times

Figure 3.6: Examples of the cells augmented by utilizing the copy-and-paste strategy. (a) Original Images. (b) Copying and pasting each cells by one time. (c) Copying and pasting each cells by three time. (d) Copying and pasting each cells by five time.

ensure that new pasted cells do not overlap with any existing target cells and is at least 5 pixels away from the image boundary. Before pasting a mitotic cell to a new location, we randomly transform it. The target rotation range involves 90°, 180° and 270°. After applying the copy-and-paste data augmentation method, the number of mitotic cells in a pathology image is increased, which in turn alleviates the low density of mitotic cells problem. By increasing the number of mitotic cells in each image, the number of matched anchors will also increase, which in turn increases the contribution of mitotic cells to loss calculations during the training phase. Additionally, we also copy and paste each cells by three and five times with the same procedure as by one time.

Chapter 4

Experiments

In this chapter, we firstly describe the ICPR MITOSIS dataset and the evaluation method. Secondly, we describe our environment settings for performing the experiments. Thirdly, we evaluate the performance of our proposed method for detecting mitotic cells on ICPR MITOSIS 2012 and 2014 dataset. Eventually, we will evaluate the performance of our model by different experimental parameters.

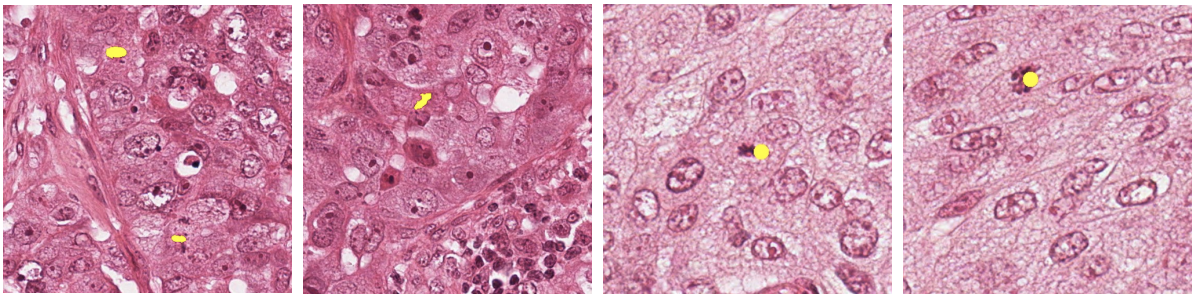
4.1 ICPR MITOSIS Dataset

In this thesis, we evaluate our lightweight YOLOv4 on two public competition dataset named ICPR MITOSIS 2012 and ICPR MITOSIS 2014. The details of each dataset are described as follows.

ICPR MITOSIS 2012: The MITOSIS 2012 dataset contains 50 histopathological images. These images are high-power fields (HPFs) which are selected from the biopsy images of five breast-cancer patients. Two scanners are used named Aperio XT (scanner A) scanner and Hamamatsu NanoZoomer (scanner H) scanner. In our experiment, images from scanner A are used. Based on the instructions, 35 HPFs are used for training, and the remaining 15 HPFs are used for testing. There are 226 mitotic cells in the training dataset, and 101 mitotic cells in the test dataset. The size of each HPF is 2084×2084 pixels, and these images are at 40X

magnification. These mitotic cells are annotated in pixel level.

ICPR MITOSIS 2014: Compared with MITOSIS 2012 dataset, the MITOSIS 2014 dataset contains more images. This dataset involves 1200 training HPFs from 16 different biopsies and 496 testing HPFs acquired from five different breast biopsies. The size of each HPF is 1539 $\times$ 1376 pixels at 40X magnification. However, only the centroid of each mitotic cell is annotated in ICPR MITOSIS 2014 dataset. Therefore, different researchers may generate different bounding boxes. We manually draw the bounding box around cells and the annotations have been verified by pathologists from the WHPG to ensure the accuracy. In our experiments, as the ground truths of the testing data are not provided by the organizers, we performed the experiments by splitting the training data into training and validation sets using the same split way mentioned in [23, 24, 2, 10] for obtaining a fair comparison, which randomly sample 80% images from the MITOSIS 2014 training data as the training set, and the remaining 20% images as the validation set. Figure 4.1 shows several examples of ICPR MITOSIS 2012 and 2014. Mitotic cells distributes in a low density. There are plenty of nuclei in each patches at



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

Figure 4.1: (a) and (b) are examples of patches in ICPR MITOSIS 2012, (c) and (d) are examples of patches in ICPR MITOSIS 2014 dataset.

40X magnification, but only a few are mitosis. It makes the mitotic cells detection task difficult due to the low occurrence rate. There are some other cells (like apoptotic cells, dense nuclei) that have similar appearances with mitosis, making it hard to filter them out as well.

4.2 Performance Evaluation Method

The performance of the proposed method is evaluated based on the number of correctly detected mitotic cells. Following by the contest criteria, a true positive is defined as a positive that exists less than 5 μm (20 pixels) from ground truth position in the MITOSIS 2012. Similarly, a true positive is defined as a positive that no more than 8 μm (32 pixels) from ground truth position in MITOSIS 2014 datasets. Therefore, based on the criteria, some measures are defined to evaluate the performance of detection. True positive (TP) is utilized to measure the number of mitotic cells which are correctly detected as mitotic cells. False positive (FP) is used to measure the number of non-mitotic cells that are incorrectly detected as mitotic cells. False negative (FN) is adopted to measure the number of ground truth mitotic cells which are not detected as mitotic cells. TP, FP and FN are used to calculate the precision, recall, and F1-Score. Eventually, F1-Score is used to determine the detection performance. The formulas are shown as follows.

$$recall = \frac{TP}{(TP + FN)} = \frac{TP}{groundtruth} \tag{4.1}$$

$$precision = \frac{TP}{(TP + FP)} = \frac{TP}{predictions} \tag{4.2}$$

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{4.3}$$

4.3 Implementation Details

We utilize and modify the public source code of YOLOv4 implemented in Darknet [1] to build a model for the mitotic cells detection task. The proposed YOLOv4 model was trained on the SharcNet Cedar cluster with Tesla P100-PCIE-12GB GPU.

[1] darknet

The size of the input image should be set to a multiple of 32 due to the 5 times of down-sampling in CSPDarknet53. Therefore, the size of a patch we crop from an ICPR MITOSIS 2012 HPF is 512×512 pixels, and the size of cropped patch is and 800×800 pixels in the ICPR MITOSIS 2014. In our experiments, rather than using the horizontal flipping, vertical flipping, translation operations mentioned in [31, 23] for augmenting data, we utilize the copy-and-paste method.

During the training process, due to the restriction of resolution and size of our samples, 16 images are loaded into the memory at once. The forward propagation is completed in 16 times by calculating 1 image each time. Eventually, the loss of the 16 images is accumulated and averaged, and the weight parameters are updated in a one-time backward propagation. In our case, the stochastic gradient descent (SGD) method was adopted for the optimization. It efficiently optimizes all the learnable parameters of the model. The hyper-parameters of SGD for the two models are set as below: learning rate 0.001, momentum 0.949 and decay 0.0005.

4.4 Evaluation

We compare our network to different current methods [23, 2, 24, 25, 31]. Table 4.2 and Table 4.3 illustrate the experiment results on ICPR MITOSIS 2012 and ICPR MITOSIS 2014 dataset. Our approach achieves a precision, recall and F1-score of 76%, 88%, 81.6%, repectively in MITOSIS 2012 dataset with the highest recall value. We also achieve a balanced precision, recall, and F1-score of 54%, 60%, and 56.8%, respectively on MITOSIS 2014 dataset. Many of the images in the IPCR MITOSIS 2014 dataset do not contain any mitosis, therefore, our proposed detection model does not perform good as in the ICPR MITOSIS 2012 dataset. In addition, it is worth noting that we consider mitosis detection as an object detection problem. Methods proposed in [23, 31] achieve high detection performance by utilizing additional CNN classifiers to filter the false positives generated from the detection model. The detection models of these methods are developed based on the R-CNN methods. However, we directly generate

Methods	ICPR MITOSIS 2012	ICPR MITOSIS 2014
Ciresan et al. [6]	31s	N.R
Li et al [23].	0.72s	0.4s
Our method	**0.5**	**0.24s**

Table 4.1: Time analysis. N.R. refers to not report.

detection results based on YOLOv4 method without utilizing any additional classifiers to filter the false positives. Table 4.4 presents the comparison results between our method and Li et al [23] method without utilizing additional classifiers. It is clearly illustrated that our method performs 7.9% better and faster than them.

The automatic mitotic cells detection system aims to provide a second-opinion for pathologists to diagnose diseases in clinical applications. Due to the large size of a single whole slide image, it is critical to have an algorithm which can detect mitotic cells both accurately and fast. The method proposed by Ciresan et al. [6] takes 31s to detect mitotic cells on an input HPF from ICPR MITOSIS 2012. The method proposed by Li et al. [23] takes 0.72s to detect mitotic cells in a HPF with a spatial dimension of 2084×2084 pixels and the RPN takes 0.68s in NVIDIA GeForce GTX TITAN X. Their method also takes 0.4s to detect mitotic cells on an HPF from the ICPR MITOSIS 2014 dataset. However, Mahmood et al. [31] did not mention the inference speed of their approach. They improve the detection performance based on the methods proposed by Li et al. with Faster RCNN and additional deep neural networks for post-processing. Therefore, the inference speed of the approach proposed by Mahmood et al. should be similar as Li et al.

Compared with these aforementioned methods, our method takes 0.5s to detect mitotic cells under Tesla P100-PCIE-12GB GPU on an input HPF from ICPR MITOSIS 2012. For a HPF with a spatial dimension of 1539×1376 pixels in ICPR MITOSIS 2014 dataset, our method takes 0.24s to detect mitotic cells. It turns out that our approach achieves competitive results with a much faster speed. The comparison is show at table 4.1. The efficiency of our method makes it more practical in practical scenario.

Methods	Precision	Recall	F1-Score
Sommer et al. [44]	51.9%	79.8%	62.9%
Malon et al. [32]	74.7%	59%	65.9%
Tashk et al. [47]	69.9%	72%	70.9%
Irshad [19]	69.8%	74%	71.8%
Ciresan et al. [6]	88%	70%	78.2%
Wang et al. [48]	84%	65%	73.5%
Paul et al. [34]	83.5%	81.1%	82.3%
Chen et al. [4]	80.4%	77.2%	78.8%
Li et al. [25]	78%	79%	78.4%
Li et al. [23]	85.4%	81.2%	83.2%
Li et al. [24]	84.6%	76.2%	80.2%
Mahmood et al. [31]	87.6%	84.1%	85.8%
Our method	**76.0%**	**88.0%**	**81.6%**

Table 4.2: Compare the literature's results on ICPR MITOSIS 2012 with ours. N.R. refers to not report.

Methods	Precision	Recall	F1-Score
Li et al [23].	N.R.	N.R.	57.2%
Cai et al. [2]	53.0%	66.0%	58.5%
Mahmood et al. [31]	84.8%	58.3%	69.1%
Our method	**54.0%**	**60.0%**	**56.8%**

Table 4.3: Compare the literature's results on ICPR MITOSIS 2014 with ours. N.R. refers to not report.

Methods	Precision	Recall	F1-Score
Li et al [23]. detection model	N.R.	N.R.	48.9%
Our method	**54.0%**	**60.0%**	**56.8%**

Table 4.4: Compare the detection method without utilizing additional classifiers in ICPR MITOSIS 2014 with ours. N.R. refers to not report.

4.5 Parameters Studies

We present the parameters studies of our proposed approach to examine the improvements caused by the different parameters in MITOSIS 2012 dataset. Besides the precision, recall, and F1-score. We test the detection speed of our model on a Tesla P100-PCIE-12GB. It is worth noting that the detection speed can be varied differently on different GPUs on the same dataset. The meanings of model abbreviations are as follows. $\mathbf{CP}_{\#}$: Each cell is copied and pasted by # time, # denotes as 1, 3 and 5. $\mathbf{YOLOv4}_{C\#_{CIoU}}$: CIoU loss used in feature layer C# where # refers to the feature layer 3, 4, 5. Same for $\mathbf{YOLOv4}_{C\#_{DIoU}}$.

Table 4.5 presents the comparison results between using traditional NMS and DIoU NMS as post-processing methods in training the YOLOv4. It takes 8s to make inference on the test dataset. It is illustrated that DIoU-NMS increases the F1-score by 1.8% with same inference speed. The result states that DIoU-NMS preserves more suitable predicted bounding boxes rather than traditional NMS.

Methods	Precision	Recall	F1-score
YOLOv4_{NMS}	74%	84%	78.6%
$\text{YOLOv4}_{NMS_{DIoU}}$	81%	80%	80.4%

Table 4.5: Compare our results with NMS and DIoU-NMS.

Table 4.6 presents the experimental results of different models. All these models are trained with DIoU-NMS but with different loss functions in two different layers. It is illustrated that these models achieve similar and better performance but detect faster than the model with 3 prediction heads in YOLOv4. This is because the sizes of mitotic cells do not change significantly as the sizes of objects in MS COCO and PASCAL VOC dataset. Therefore, we do not need 3 prediction heads. Furthermore, the receptive field of each feature maps can only cover a limited scale range, which resulting in poor performance if the objects' scales mismatches with the receptive field. In the original YOLOv4 with 3 prediction heads, the anchors are pre-defined on multiple levels, and the ground-truth boxes generate positive anchors in feature

Methods	Precision	Recall	F1-score
YOLOv4$_{C3_{CIoU},C4_{CIoU}}$	79%	80%	79.5%
YOLOv4$_{C3_{DIoU},C4_{CIoU}}$	80%	82%	80.9%
YOLOv4$_{C3_{DIoU},C4_{DIoU}}$	81%	79%	80%
YOLOv4$_{C3_{DIoU},C5_{DIoU}}$	76%	84%	79.8%
YOLOv4$_{C3_{DIoU},C5_{CIoU}}$	81%	76%	78.4%
YOLOv4$_{C3_{CIoU},C5_{CIoU}}$	78%	81%	79.5%
YOLOv4$_{C4_{DIoU},C5_{CIoU}}$	**76%**	**88%**	**81.6%**
YOLOv4$_{C4_{DIoU},C5_{DIoU}}$	73%	83 %	77.7%
YOLOv4$_{C4_{CIoU},C5_{CIoU}}$	80%	83%	81.5%

Table 4.6: Compare our results with using 2 Prediction heads, 8 anchors and copying cells and paste them just once.

levels corresponding to their scales. In our case, we have less prediction heads, therefore, we increase the number of anchors in each scales from 3 to 4. We only generate predictions on two prediction heads, the lightweight model infers faster than the model with three prediction heads.

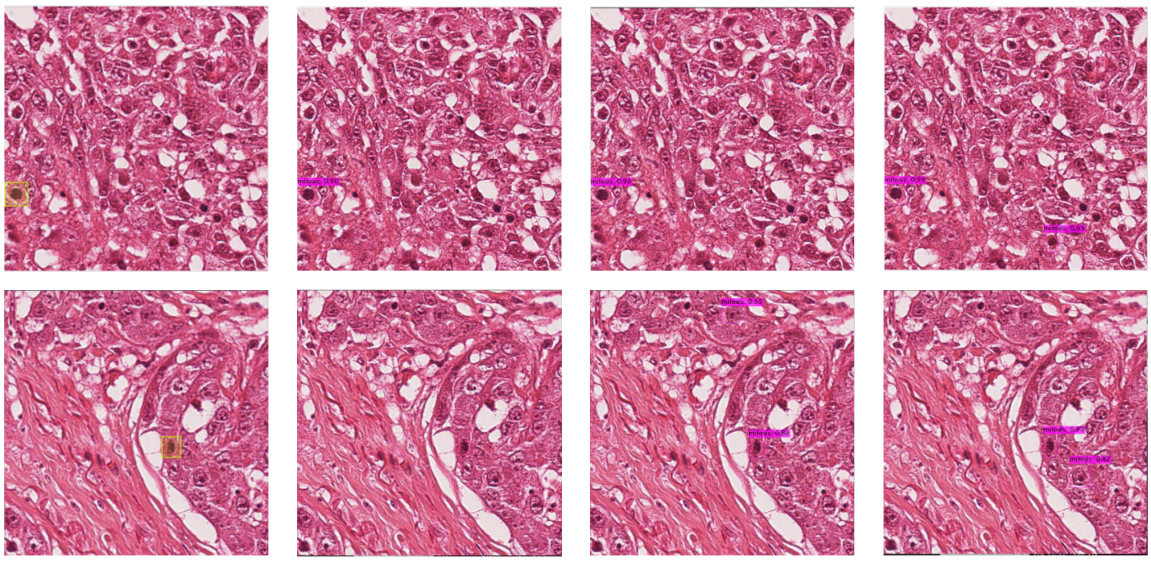
In addition, these models take 7s to make inference on the test dataset which is faster than the model with 3 YOLO heads. These models do not need to contribute to all the YOLO heads in Figure 3.1. it is illustrated that the models trained with DIoU loss and CIoU loss together perform better than the models trained with DIoU loss and CIoU loss separately. And the model trained with CIoU loss used only for C5 and DIoU loss used in C4 performs the best with the highest F1-score. This experimental result verify our assumptions and ideas that the $\alpha \times v$ term in the CIoU loss does not contribute much for small and medium objects, and the $\alpha \times v$ term works better for the large receptive field feature map.

We also utilize different copy-and-paste approaches to pre-process the training dataset. By copying and pasting small mitotic cells in each image by different times, the number of matched anchors will also increase, which in turn increases the contribution of small cells to loss calculations during the training phase. Our experimental results demonstrate that the models with copy-and-paste method by three times perform the best as shown in the table 4.7. Figure 4.2

Methods	Precision	Recall	F1-score
YOLOv4$_{CP_1}$	73%	85%	78.5%
YOLOv4$_{CP_3}$	76%	88%	81.6%
YOLOv4$_{CP_5}$	75%	77%	76%

Table 4.7: Results of adopting different copy-and-paste approaches in mitotic cells detection task.

demonstrates some prediction results of our model trained by copying and pasting mitotic cells in different times.



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

Figure 4.2: Examples of comparison results among models trained with and without using copy-and-paste methods. (a) Ground-Truth. (b) Results from model trained with CP_1. (c) Results from model trained with CP_3. (d)Results from model trained with CP_5.

Chapter 5

Fast Learning with A Few Images

A limited number of annotated samples in a practical scenario will lead to a poor performance of detection model. Therefore, we design and implement a fast learning pipeline to facilitate the detection model of weak performance which is issued by less number of labeled images. In this chapter, we evaluate the performance of the proposed methods for mitosis detection on the WHPG dataset. We also describe our online progressive "correction and relearning" pipeline in detail. This pipeline aims to correct the wrong detection results and re-training a new model to generate better predictions for future samples.

5.1 WHPG Dataset

Besides the mitosis in breast cancer, pathologists are also interested in other regions, for example, glioma, melanoma, and meningioma. In this work, we collaborate with the pathologists from WHPG to build a computational tool. This tool aims to automatically count the number of mitosis in a Whole Slide Image (WSI). The WSI, also known as virtual microscopy, refers to scanning a complete microscope slide and creating a single high-resolution digital file. It is one type of computerized medical imaging, and is the most recent imaging modality being used by pathology departments worldwide. However, the analysis of digital WSIs remains challenging because of their extremely high spatial resolution compared with other medical imaging

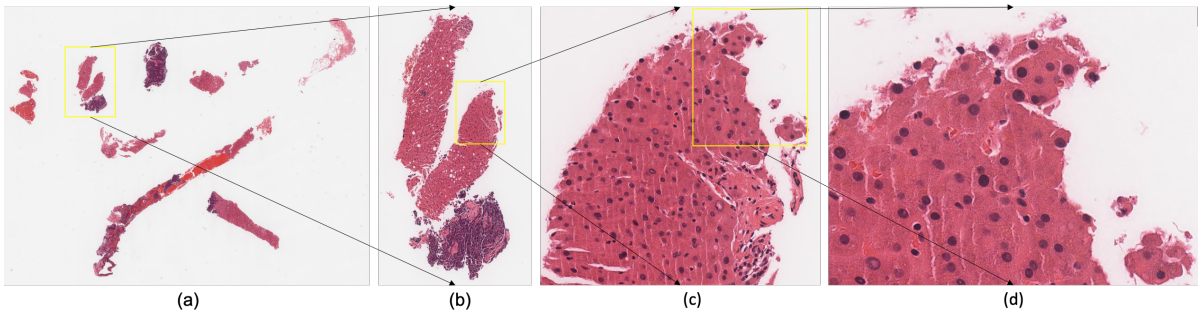


Figure 5.1: An example of a Whole Slide Image. (a) is the original image. (b) is at 4X magnification. (c) is at 20X magnification. (d) is at 40X magnification.

modalities. The size of a WSI is approximately 100000 x 100000 pixels. It is extremely difficult to train a model based on the original WSIs. The figure 5.1 demonstrates an example of WSI at different magnifications.

Generally, the size of each WSI is extremely bigger than usual MS COCO or PASCAL VOC dataset. It is hard to make annotations on an original WSIs due to its large size. Even for pathologists, they roughly determine the categories of each cells under at least 4X magnification. Therefore, in order to make distinct features, pathologists commonly make annotations under 40X magnification on different regions of each WSI. Restricted by computing limitations, most histopathology studies start by cropping patches from a WSI, followed by an object detection model to reveal the final diagnosis. These pipeline have yielded successful results in mitosis detection tasks [23, 24, 3, 31].

Considering on the resources and the cost on annotating images in a real word scenario, the size of each image in WHPG is 800×800 pixels at 40X magnification, which is smaller than the size of images in ICPR MITOSIS dataset. Figure 5.2 illustrates several examples with annotations. The WHPG dataset contains 256 images at 40X magnification. These images are cropped from different WSIs from glioma, melanoma, meningioma, and gastrointestinal neurodendocrine tumor cases. 385 mitosis are marked in this dataset, and the annotations on bounding boxes are provided and verified by two pathologists from the WHPG in one month. We randomly sampled 90% images from the dataset as the training dataset, and the remaining 10% imaged are used as the test dataset.

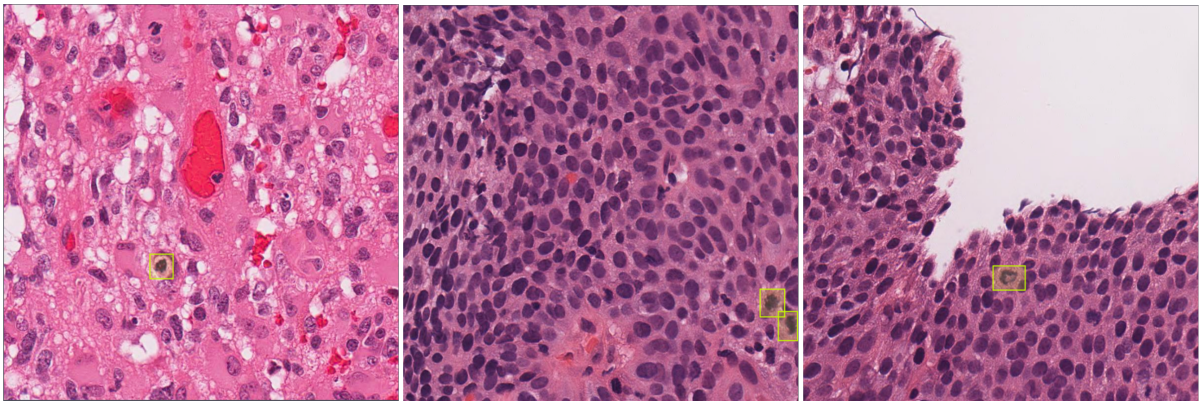


Figure 5.2: Examples of images in WHPG dataset with annotations.

Compared with the ICPR MITOSIS dataset, we do not need to crop the images to small patches. Therefore, WHPG dataset has less number of samples for training than the ICPR MITOSIS dataset. This makes the mitosis detection more difficult on the WHPG dataset.

5.2 Detection Procedure on a Whole Slide Image

mitosis count is a critical biomarker in breast cancer grading system, and can also be utilized to examine the grade of other cancers. In a practical scenario, pathologists usually examine mitosis on several ROIs on a WSI. However, the size of a WSI is extremely big, and the sizes of ROIs are sometimes relatively big as well. It is difficult to directly generate detection results.

Therefore, in order to make predictions on several ROIs selected by pathologists of a WSI, we firstly crop these ROIs from the WSI, and magnified them into 40X magnification. Second, each of the ROIs is further divided into several 800×800 patches and our model is applied to generate predictions on these patches. If the sizes of some ROIs are not a multiple of 800×800, for instace, 1500×1500, we pad the value zero around the edges of the RoI to make it into a multiple of 800. Then, we further divide this ROI into 4 patches with same size of 800×800. Eventually, all of the detected results are stitched back to reveal all of the detection results in the original WSI. In practical scenario, it cost half four minutes to generate predictions on a WSI. Dividing a WSI and stitching patches back cost most of the time. This is an inevitable

overhead. The general procedure is demonstrated in the figure 5.3.

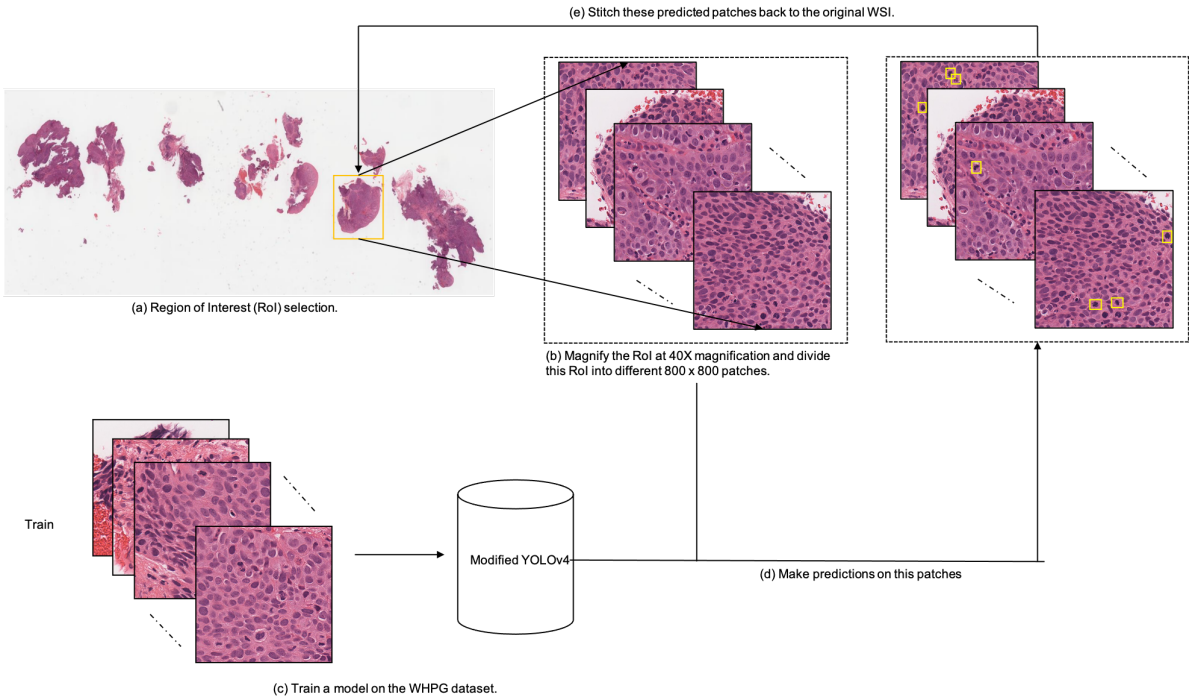


Figure 5.3: Whole procedure for mitosis detection task in practical scenario. (a) Pathologists select RoI(s). (b) We magnify and divide the RoI(s) into different patches with same size at 40X magnification. (c) We use the pre-traiend model to initialize the training of WHPG dataset. (d) We generate predictions on the divided patches. (e) We stitch these patches with predictions together to reveal the RoI(s) on a WSI.

5.3 Training

The WHPG dataset contains less number of samples for training compared with the ICPR MITOSIS dataset. Generally, a complicated model trained on a small size dataset will produce poor generalization on the test dataset. YOLOv4 is a complicated deep neural network. In order to perform better on the WHPG dataset, we use the models trained on ICPR MITOSIS 2012 and 2014 as pre-trained models for the WHPG dataset. Due to the similarity between these two datasets, the pre-trained model has learned effective features of mitosis. We have retained the architecture and the weights of the pre-trained model, and used them to initialize and retrain a new model on the WHPG dataset.

5.4 Evaluation

We use the precision, recall and F1-score to evaluate the detection performance on the WHPG dataset. Table 5.1 illustrates that the model trained with the pre-trained model from the ICPR MTOSIS 2014 dataset performs better. The precision, recall, and F1-score of 79%, 68%, and 73%, are achieved respectively by using the pre-trained model from 2014[1]. Compared with the training from scratch, the value of F1-score is increased by 16%. All of the programs are trained and tested on the SharcNet Cedar cluster Tesla P100-PCIE-12GB as well.

Methods	Precision	Recall	F1-score
$\text{Train}_{scratch}$	51%	65%	57%
$\text{Train}_{YOLOv4_{2012}}$	80%	59%	68%
$\text{Train}_{YOLOv4_{2014}}$	79%	68%	73%

Table 5.1: Results of training from scratch and training with pre-trained model.

5.5 Correction and Relearning Pipeline

Both deep learning based methods and manual examinations by pathologists make mistakes on predicting mitosis in unseen images. One possible reason is that the current training images do not cover all of the types of mitosis. False positives and false negatives are generated due to the limited training images. However, it is time-consuming to annotate enough mitosis.

With pathologists constantly examining new images, the model trained on existing dataset is no longer able to provide satisfactory results. It is beneficial for pathologists to eliminate false positives and re-annotate false negatives during the daily examinations. These modifications in the annotations on images can be used to expand the existing dataset. The new training dataset with more images can be used to train an new model to improve the detection performance. These relabeled and deleted annotations on cells should be correctly appeared and not appeared respectively in the future detection. Therefore, it is critical to develop a "correc-

[1]Due to the small limited WHPG dataset, we did not use the validation dataset when we are training.

tion and relearning" progressive pipeline to continuously learn from mistakes and learn new features from new images.

Pathologists often examine new WSIs or different ROIs in a WSI. We consider the continuous flow of new examination samples as stream data. Our model firstly predicts detection results on these new samples. Afterward, the "correction" procedure is utilized to modify the false positives and false negatives. These modified samples are accumulated with the existing images together in the training of a new model after each modified mitosis is copied and pasted once. In the "relearning" procedure, we retrain different new models based on the models trained from ICPR MITOSIS 2014 dataset, the new model with the highest F1-score is chosen for pathologists to generate further detections. Our model can be updated progressively with more training samples under this "correction and relearning" pipeline. The following algorithm pseudo-code demonstrates the pipeline. θ refers to the number of images being corrected, in our case, it is 5.

Algorithm 1: Correction and Relearning

Input : Images **I**
Output: Predicted Images **PIs**

$\text{Models}_{pretrain} \leftarrow \text{Models}_{MITOSIS\,2014}$
$\text{I}_{init} \leftarrow \mathbf{0}$
$\text{Models} \leftarrow \text{Models}_{pretrain}$
while $Num_I \neq 0$ **do**
    $\text{PIs} \leftarrow \text{Models(I)}$
    /* Use the models to predict images. */
    $\text{PIs}_{re-annotated} \leftarrow$ Re-annotating from experts
    $\text{PIs}_{re-annotated} \leftarrow \text{Copy-and-Paste}(\text{PIs}_{re-annotated})$
    $\text{I}_{init} \leftarrow I_{init} + \text{PIs}_{re-annotated}$
    /* Accumulate re-annotated images to current dataset. */
    if $Num_{PIs_{re-annotated}} == \theta$ **then**
        $\text{Models} \leftarrow \text{Models}_{pretrain}(\text{I}_{init})$
        /* Relearn the new dataset to retrain new models. The best new model is selected with the highest F1-score */
    end
end
return *PIs*

In a practical scenario, we use the WHPG dataset to simulate the data stream. In order to

evaluate the performance of using this progressive pipeline, we use 10% images from WHPG as the ground truth. In each step, we feed the detection model with only 5 images to generate predictions. Figure 5.4 demonstrates the progressive results. It is illustrated that with the number of images increasing continuously, the new model after relearning performs a better result. However, the training time is getting longer due to the increased number of training samples. Figure 5.5 illustrates the validity of relearning pipeline. It is demonstrated that false negatives can be detected after relearning.

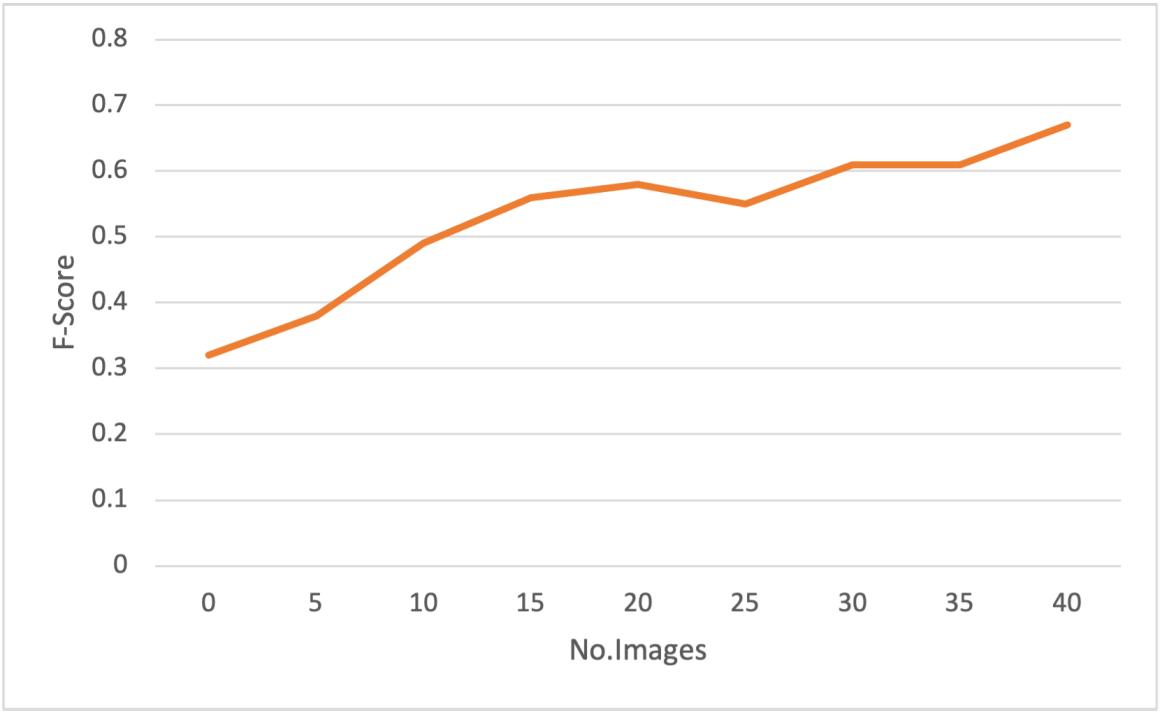


Figure 5.4: progressive result.

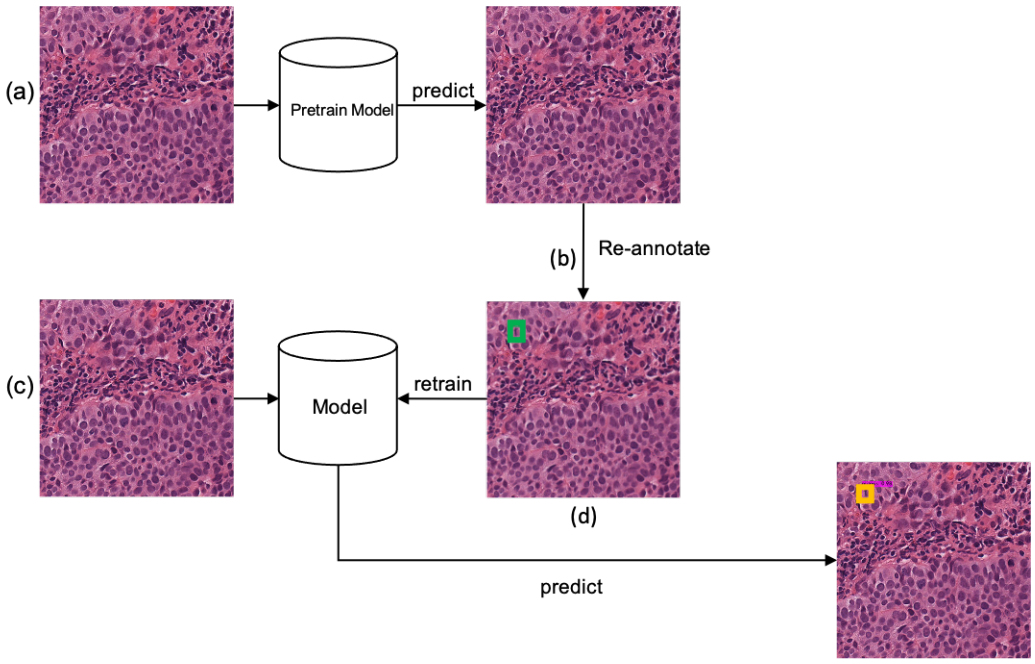


Figure 5.5: An overview of "correction and relearning" pipeline.(a) Pathologists use our pre-trained model to predict images. (b) Pathologists make corrections on the predicted results. These modified samples are accumulated with the existing images together in the training of a new model after each modified mitosis is copied and pasted once. (c)-(d) Correct results will be predicted based on the retrained model.

Chapter 6

Conclusion and Further Work

In this chapter, we firstly conclude our work presented in this thesis. Lastly, we discuss our future plans.

6.1 Conclusion

Currently, most of mitotic cells detection works are performed by pathologists manually under high-resolution microscopes, which is challenging and time-consuming. Previous work have used artificial intelligence approaches to automatically detect the mitotic cells. It has been exhibited that these previous works do not satisfy the inference speed due to utilizing two-stage object detection frameworks and additional binary classifiers for post-processing. In order to release the workload, in this work, we modify and present a lightweight mitotic cells detection model based on YOLOv4. This modified model is developed based on reducing the number of predicting heads, utilizing different loss functions for different sizes of mitotic cells, generating more accurate anchors and employing DIoU-NMS method to suppress the redundant predict bounding boxes. We also augment the cell density by adopting copy-and-paste approach.

In the literature review section, we describe the previous works developed in the mitotic cells detection task and some popular object detection models in computer vision domain. We also explain the disadvantages of each method. In order to overcome these disadvantages,

we introduce the YOLO object detection framework. Eventually, we present different data augmentation methods utilized in computer vision domain, and describe the copy-and-paste augmentation methods used in object detection task. In the experiments section, the results of our approach on the public online competition dataset named ICPR MITOSIS 2012 and ICPR MITOSIS 2014 are demonstrated. It is illustrated that our proposed approach achieves competitive result on ICPR MITOSIS 2012 dataset and on ICPR MITOSIS 2014 dataset but with faster inference speed. We also apply the model trained from ICPR MITOSIS 2012 and 2014 on the WHPG dataset, it states that our approach performs better than the model training from scratch by 16% by using the pre-trained model from 2014 dataset.

In addition, a few number of annotated samples in a practical scenario will lead to a poor performance of detection model. Therefore, we propose a progressive "correction and relearning" pipeline for pathologists. This pipeline aims to continuously learn from the mistakes and learn new knowledge. Satisfactory results can be achieved through annotating or re-annotating only a small number of slides, and the detection performance is boosted while increasing the training samples.

In conclusion, we propose an efficient and lightweight model to faster and more accurately detect mitotic cells based on YOLOv4. We compare our results with previous works in the mitotic cell detection task on the ICPR MITOSIS 2012 and ICPR MITOSIS 2014 dataset. In practical scenario, we use the pre-trained model learned from ICPR MITOSIS 2014 dataset to train the model on WHPG dataset. The effectiveness of our method is proved by the experimental results.

6.2 Future work

We believe that the succeed in computer vision domain can be adopted in pathological cell detection domain, we intend to continuously improve the detection accuracy and the decrease the computational cost. In addition, some changes are needed due to the difficulties in obtain-

ing enough annotated training dataset. The proposed "correction and relearning" pipeline can facilitate the detection model of weak performance which is issued by less number of labeled images. However, it is redundant and time-consuming to train the model from scratch due to loading the accumulated dataset each time. A smarter way is to retrain a model only based on the new samples rather than retraining a model based on the entire dataset. Therefore, in the future, we plan to utilize the concept of incremental learning to retrain a model only based on the new samples and produce a better detection performance. In addition, current methods achieves high detection performance by utilizing additional classifiers to filter the false positives generated from the detection model. In this thesis, we only explore the performance by using YOLOv4 to directly detect mitosis without using additional classifiers. Therefore, in the future, we plan to adopt additional classifier to filter the false positives which can boost the detection performance.

Bibliography

[1] Alexey Bochkovskiy, Chien-Yao Wang, and Hong-Yuan Mark Liao. Yolov4: Optimal speed and accuracy of object detection. *arXiv preprint arXiv:2004.10934*, 2020.

[2] De Cai, Xianhe Sun, Niyun Zhou, Xiao Han, and Jianhua Yao. Efficient mitosis detection in breast cancer histology images by rcnn. In *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*, pages 919–922. IEEE, 2019.

[3] Zhaowei Cai and Nuno Vasconcelos. Cascade r-cnn: Delving into high quality object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 6154–6162, 2018.

[4] Hao Chen, Qi Dou, Xi Wang, Jing Qin, and Pheng Heng. Mitosis detection in breast cancer histology images via deep cascaded networks. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 30, 2016.

[5] François Chollet. Xception: Deep learning with depthwise separable convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1251–1258, 2017.

[6] Dan C Cireşan, Alessandro Giusti, Luca M Gambardella, and Jürgen Schmidhuber. Mitosis detection in breast cancer histology images with deep neural networks. In *International conference on medical image computing and computer-assisted intervention*, pages 411–418. Springer, 2013.

[7] Ekin D Cubuk, Barret Zoph, Dandelion Mane, Vijay Vasudevan, and Quoc V Le. Autoaugment: Learning augmentation policies from data. *arXiv preprint arXiv:1805.09501*, 2018.

[8] Ekin D Cubuk, Barret Zoph, Jonathon Shlens, and Quoc V Le. Randaugment: Practical automated data augmentation with a reduced search space. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 702–703, 2020.

[9] Navneet Dalal and Bill Triggs. Histograms of oriented gradients for human detection. In *2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05)*, volume 1, pages 886–893. Ieee, 2005.

[10] Veena Dodballapur, Yang Song, Heng Huang, Mei Chen, Wojciech Chrzanowski, and Weidong Cai. Mask-driven mitosis detection in histopathology images. In *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*, pages 1855–1859. IEEE, 2019.

[11] Debidatta Dwibedi, Ishan Misra, and Martial Hebert. Cut, paste and learn: Surprisingly easy synthesis for instance detection. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 1301–1310, 2017.

[12] Golnaz Ghiasi, Yin Cui, Aravind Srinivas, Rui Qian, Tsung-Yi Lin, Ekin D Cubuk, Quoc V Le, and Barret Zoph. Simple copy-paste is a strong data augmentation method for instance segmentation. *arXiv preprint arXiv:2012.07177*, 2020.

[13] Ross Girshick, Jeff Donahue, Trevor Darrell, and Jitendra Malik. Rich feature hierarchies for accurate object detection and semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 580–587, 2014.

[14] Alaa Ali Hameed, Bekir Karlik, and Mohammad Shukri Salman. Back-propagation algorithm with variable adaptive momentum. *Knowledge-Based Systems*, 114:79–87, 2016.

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

[16] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Spatial pyramid pooling in deep convolutional networks for visual recognition. *IEEE transactions on pattern analysis and machine intelligence*, 37(9):1904–1916, 2015.

[17] 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.

[18] Michael Held, Michael HA Schmitz, Bernd Fischer, Thomas Walter, Beate Neumann, Michael H Olma, Matthias Peter, Jan Ellenberg, and Daniel W Gerlich. Cellcognition: time-resolved phenotype annotation in high-throughput live cell imaging. *Nature methods*, 7(9):747–754, 2010.

[19] Humayun Irshad. Automated mitosis detection in histopathology using morphological and multi-channel statistics features. *Journal of pathology informatics*, 4, 2013.

[20] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, 25:1097–1105, 2012.

[21] Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998.

[22] Haijun Lei, Shaomin Liu, Hai Xie, Jong Yih Kuo, and Baiying Lei. An improved object detection method for mitosis detection. In *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 130–133. IEEE, 2019.

[23] Chao Li, Xinggang Wang, Wenyu Liu, and Longin Jan Latecki. Deepmitosis: Mitosis detection via deep detection, verification and segmentation networks. *Medical image analysis*, 45:121–133, 2018.

[24] Chao Li, Xinggang Wang, Wenyu Liu, Longin Jan Latecki, Bo Wang, and Junzhou Huang. Weakly supervised mitosis detection in breast histopathology images using concentric loss. *Medical image analysis*, 53:165–178, 2019.

[25] Yuguang Li, Ezgi Mercan, Stevan Knezevitch, Joann G Elmore, and Linda G Shapiro. Efficient and accurate mitosis detection-a lightweight rcnn approach. In *ICPRAM*, pages 69–77, 2018.

[26] Rainer Lienhart and Jochen Maydt. An extended set of haar-like features for rapid object detection. In *Proceedings. international conference on image processing*, volume 1, pages I–I. IEEE, 2002.

[27] Tsung-Yi Lin, Piotr Dollár, Ross Girshick, Kaiming He, Bharath Hariharan, and Serge Belongie. Feature pyramid networks for object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2117–2125, 2017.

[28] Shu Liu, Lu Qi, Haifang Qin, Jianping Shi, and Jiaya Jia. Path aggregation network for instance segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 8759–8768, 2018.

[29] David G Lowe. Distinctive image features from scale-invariant keypoints. *International journal of computer vision*, 60(2):91–110, 2004.

[30] Gjorgji Madzarov, Dejan Gjorgjevikj, and Ivan Chorbev. A multi-class svm classifier utilizing binary decision tree. *Informatica*, 33(2), 2009.

[31] Tahir Mahmood, Muhammad Arsalan, Muhammad Owais, Min Beom Lee, and Kang Ryoung Park. Artificial intelligence-based mitosis detection in breast cancer histopathology images using faster r-cnn and deep cnns. *Journal of clinical medicine*, 9(3):749, 2020.

[32] Christopher D Malon and Eric Cosatto. Classification of mitotic figures with convolutional neural networks and seeded blob features. *Journal of pathology informatics*, 4, 2013.

[33] Xipeng Pan, Yinghua Lu, Rushi Lan, Zhenbing Liu, Zujun Qin, Huadeng Wang, and Zaiyi Liu. Mitosis detection techniques in h&e stained breast cancer pathological images: A comprehensive review. *Computers & Electrical Engineering*, 91:107038, 2021.

[34] Angshuman Paul, Anisha Dey, Dipti Prasad Mukherjee, Jayanthi Sivaswamy, and Vijaya Tourani. Regenerative random forest with automatic feature selection to detect mitosis in histopathological breast cancer images. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 94–102. Springer, 2015.

[35] Jianghao Rao and Jianlin Zhang. Cut and paste: Generate artificial labels for object detection. In *Proceedings of the International Conference on Video and Image Processing*, pages 29–33, 2017.

[36] Joseph Redmon, Santosh Divvala, Ross Girshick, and Ali Farhadi. You only look once: Unified, real-time object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 779–788, 2016.

[37] Joseph Redmon and Ali Farhadi. Yolo9000: better, faster, stronger. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 7263–7271, 2017.

[38] Joseph Redmon and Ali Farhadi. Yolov3: An incremental improvement. *arXiv preprint arXiv:1804.02767*, 2018.

[39] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. Faster r-cnn: Towards real-time object detection with region proposal networks. *arXiv preprint arXiv:1506.01497*, 2015.

[40] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, et al. Imagenet large scale visual recognition challenge. *International journal of computer vision*, 115(3):211–252, 2015.

[41] Abhronil Sengupta, Yuting Ye, Robert Wang, Chiao Liu, and Kaushik Roy. Going deeper in spiking neural networks: Vgg and residual architectures. *Frontiers in neuroscience*, 13:95, 2019.

[42] Pierre Sermanet, David Eigen, Xiang Zhang, Michaël Mathieu, Rob Fergus, and Yann LeCun. Overfeat: Integrated recognition, localization and detection using convolutional networks. *arXiv preprint arXiv:1312.6229*, 2013.

[43] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.

[44] Christoph Sommer, Luca Fiaschi, Fred A Hamprecht, and Daniel W Gerlich. Learning-based mitotic cell detection in histopathological images. In *Proceedings of the 21st International Conference on Pattern Recognition (ICPR2012)*, pages 2306–2309. IEEE, 2012.

[45] Christoph Sommer, Christoph Straehle, Ullrich Koethe, and Fred A Hamprecht. Ilastik: Interactive learning and segmentation toolkit. In *2011 IEEE international symposium on biomedical imaging: From nano to macro*, pages 230–233. IEEE, 2011.

[46] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. Going deeper with convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1–9, 2015.

[47] Ashkan Tashk, Mohammad Sadegh Helfroush, Habibollah Danyali, and Mojgan Akbarzadeh. An automatic mitosis detection method for breast cancer histopathology slide images based on objective and pixel-wise textural features classification. In *The 5th conference on information and knowledge technology*, pages 406–410. IEEE, 2013.

[48] Haibo Wang, Angel Cruz-Roa, Ajay Basavanhally, Hannah Gilmore, Natalie Shih, Mike Feldman, John Tomaszewski, Fabio Gonzalez, and Anant Madabhushi. Cascaded ensemble of convolutional neural networks and handcrafted features for mitosis detection. In *Medical Imaging 2014: Digital Pathology*, volume 9041, page 90410B. International Society for Optics and Photonics, 2014.

[49] Sangdoo Yun, Dongyoon Han, Seong Joon Oh, Sanghyuk Chun, Junsuk Choe, and Youngjoon Yoo. Cutmix: Regularization strategy to train strong classifiers with localizable features. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 6023–6032, 2019.

[50] Hongyi Zhang, Moustapha Cisse, Yann N Dauphin, and David Lopez-Paz. mixup: Beyond empirical risk minimization. *arXiv preprint arXiv:1710.09412*, 2017.

[51] Zhi Zhang, Tong He, Hang Zhang, Zhongyue Zhang, Junyuan Xie, and Mu Li. Bag of freebies for training object detection neural networks. *arXiv preprint arXiv:1902.04103*, 2019.

[52] Zhaohui Zheng, Ping Wang, Wei Liu, Jinze Li, Rongguang Ye, and Dongwei Ren. Distance-iou loss: Faster and better learning for bounding box regression. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 12993–13000, 2020.

Curriculum Vitae

Name:	Jinhang Zhang
Post-Secondary Education and Degrees:	University of Nottingham Nottingham, UK 2012 - 2016 B.S. University College London London, UK 2016- 2017 M.S.
Honours and Awards:	Western Graduate Research Scholarships(WGRS) 2019-2020
Related Work Experience:	Teaching Assistant and Research Assistant The University of Western Ontario 2019 - 2020