DeepMitosis: Mitosis Detection via Deep Detection, Verification and Segmentation Networks

Chao Li[a], Xinggang Wang[a,*], Wenyu Liu[a], Longin Jan Latecki[b]

[a] *School of Electronics Information and Communications, Huazhong University of Science and Technology, Wuhan, P.R. China.*
[b] *CIS Dept., Temple University, 1805 N. Broad St. Philadelphia, PA 19122, USA.*

Abstract

Mitotic count is a critical predictor of tumor aggressiveness in the diagnosis of breast cancer. Nowadays mitosis counting is mainly performed manually by pathologists, which is extremely arduous and time-consuming. Hence it is very necessary to develop automatic mitosis detection methods. In this paper, we propose an accurate method for detecting and counting the mitotic cells from histopathological slides using a novel multi-stage deep learning framework. Our method consists of a deep segmentation network for generating mitosis region when only a weak label is given (i.e., only the centroid location of mitosis is labeled), a carefully designed deep detection network using contextual region information to localize mitosis, and a deep verification network for improving detection accuracy by removing false positives. We validate the proposed deep learning method in the widely used Mitosis Detection in Breast Cancer Histological Images (MITOS dataset). Experimental results show that we can achieve the highest F-measure on the MITOSIS dataset from ICPR 2012 grand challenge merely by the deep detection network. For the ICPR 2014 MITOSIS dataset which only provides the centroid location of mitosis, we employ the segmentation network to perform semantic segmentation and estimate the bounding box annotation of mitosis for training the deep detection model. And

*Corresponding author.
Email addresses: chaol@hust.edu.cn (Chao Li), xgwang@hust.edu.cn (Xinggang Wang), liuwy@hust.edu.cn (Wenyu Liu), latecki@temple.edu (Longin Jan Latecki)

the verification model is applied to eliminate some false positives produced by the detection model. Fusing scores of the detection and verification models, we achieve the state-of-the-art results. Moreover, our method is very fast with GPU computing, which makes it feasible for clinical practice.

Keywords: Mitosis detection, Faster R-CNN, Fully convolutional network, Breast cancer grading

1. Introduction

According to the Nottingham Grading System (Elston & Ellis, 1991), there are three morphological features on Hematoxylin and Eosin (H&E) stained slides that are important for breast cancer grading. They are mitotic count, tubule formation, and nuclear pleomorphism. Among the three indicators, mitotic count is the most critical one. Pathologists usually search for mitosis and count their number in high-power fields (HPFs) manually. It is time-consuming and difficult due to the large number of HPFs in a single whole slide and the high variation in the appearance of mitotic cells. Besides, the judgment of mitotic cell is very subjective, and it is hard to reach a consensus on mitotic count among pathologists. Thus developing methods for detecting mitosis automatically is very essential, which will not only save a lot of time, manpower and material resources but also improve the reliability of pathological diagnosis.

Mitosis detection from the H&E stained histopathological images is hard due to some challenges. First, the appearance of mitosis varies in a wide range. Mitosis has four phases which are called prophase, metaphase, anaphase and telophase. The shapes and structures of cells in different phases are very diverse as shown in Fig. 1 (a)-(c). In the telophase stage, the nucleus of a cell has split into two parts, but they should still be counted as one cell because they are not yet separated completely from each other. Second, mitotic cells are considerably less than non-mitotic cells. The low probability of emergence makes the detection more difficult. Third, there are some other cells (like apoptotic cells, dense nuclei) that have a similar appearance with mitosis, making it hard

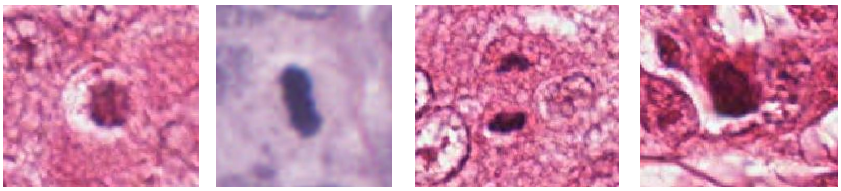

Figure 1: Examples of mitotic cells and non-mitotic cell. (a) and (b) show the prophase and metaphase of mitosis, respectively. (c) shows the appearance of a mitosis in telophase stage, which should be still counted as one cell though it has two distinct nuclei. (d) shows a non-mitotic cell having a similar appearance with mitotic cells.

to filter them out.

Recently, many automatic mitosis detection methods have been proposed. This phenomenon owes to some mitosis detection contests including the 2012 ICPR mitosis detection contest (Roux et al., 2013), the AMIDA13 contest at MICCAI 2013 (Veta et al., 2015), and the 2014 ICPR MITOS-ATYPIA challenge (MITOS-ATYPIA-14, 2014). Most of the appeared methods use handcrafted features to model the appearance of mitosis for detection. Some methods (Sommer et al., 2012; Irshad et al., 2013; Khan et al., 2012; Veta et al., 2013; Tek et al., 2013) design different sorts of statistical, morphological and textural features to capture characteristics of mitosis explicitly. However, due to the diverse and complex shapes of mitosis and the existence of confounding cells, it is hard to manually design meaningful and discriminative features to distinguish mitosis from non-mitotic cells effectively.

Since the remarkable work of Alex et al. (Krizhevsky et al., 2012) in ILSVRC 2012 (Russakovsky et al., 2015a), convolutional neural networks (CNN) have revolutionized the world of computer vision. The methods based on CNN have set up new records in many vision tasks, such as image classification (He et al., 2016), object detection (Ren et al., 2015) and semantic segmentation (Long et al., 2015). In biomedical analysis field, CNN based methods also yield ex-

cellent performance in several tasks, for instance, the segmentation of neuronal membranes in microscopy images (Ciresan et al., 2012) and the analysis of developing embryos from videos (Ning et al., 2005). Hence some CNN based approaches (Malon et al., 2013; Cireşan et al., 2013; Wang et al., 2014; Chen et al., 2016a,b) have been proposed to detect mitosis. They utilize features learned from data automatically, and the convolutional features are more efficacious than the handcrafted features.

Current deep networks based mitosis detection methods can be divided into two folds: 1) Considering mitosis detection as a classification problem (Cireşan et al., 2013), it classifies image patches using a plain CNN. This strategy could be regarded as a sliding-window-based mitosis detection method, which is very slow. 2) Considering mitosis detection problem as a semantic segmentation problem (Chen et al., 2016b), it infers pixel-level label of mitosis using fully convolutional networks (FCN), which ignores region information and is hard to deal with the weak labels, e.g., the 2014 MITOSIS dataset only labels the center of mitosis. Thus, we argue that considering the mitosis detection problem as an object detection problem makes more sense. We propose a method named DeepMitosis which uses deep detection network to solve the mitosis detection problem. Meanwhile, to the best of our knowledge, this is the first paper that utilizes deep detection method for the mitosis detection problem.

The region-based ConvNets detection methods (Girshick et al., 2014; He et al., 2014; Girshick, 2015; Ren et al., 2015) are very prevalent in object detection field. Among them, Faster R-CNN (Ren et al., 2015) uses a fully convolutional Region Proposal Network (RPN) to generate proposals and then applies a region-based classification network to classify these proposals. It achieves excellent accuracy on the PASCAL VOC detection benchmarks (Everingham et al., 2007). We adapt the deep detection model Faster R-CNN to the mitosis detection task. An overview of our proposed DeepMitosis system is illustrated in Fig. 2. It consists of three components: a deep segmentation model based on FCN for producing estimated bounding box labels, a deep detection model based on Faster R-CNN for localizing mitosis, and a deep verification model

based on ResNet for classifying the detection patches further to improve the accuracy. The training of deep detector requires the bounding box label, and we can train it on 2012 MITOSIS dataset directly since the dataset has already given the label of each pixel. However, the 2014 MITOSIS dataset only labels the centroid of mitosis, thus we need to estimate the bounding box annotations before training the mitosis detector. Inspired by the success of FCN in semantic segmentation in natural images, we take advantage of a FCN model trained on 2012 MITOSIS dataset, to perform semantic segmentation on the weakly annotated 2014 MITOSIS dataset. Combining the segmentation results with the original centroid labels, we can infer a bounding box label for every mitotic cell. The predicted box labels are then utilized to train the deep detection model. As shown in Fig. 2 (b), the detection process has two stages. Firstly we run the deep detection model on a histology image to produce detection results, and then these detected image patches are fed into the deep verification model for further refinement. The verification model is a ResNet (He et al., 2016), which has 50 layers with short-cut connections and is a powerful image classification network. The final prediction is a weighted sum of the predictions from the detection model and the verification model, which is better than either of the both.

In summary, we mainly have three contributions in this paper: (1) We design an architecture to estimate mitosis count automatically. The core of the proposed architecture is a tailored Faster R-CNN; Faster R-CNN is originally proposed for object detection in natural images. We refine the general object detection framework to medical images and achieve state-of-the-art performance on two mitosis benchmark datasets. To our best knowledge, this is the first work to apply Faster R-CNN on mitosis detection. (2) We train a deep segmentation network to estimate the region of mitosis. With the estimated mitotic region, we infer the bounding box labels and use them to train a deep detection model. Experimental results demonstrate that the inferred bounding box labels improve the performance compared to a simple square bounding box. (3) We adopt a classification model to further verify the detection results of the deep

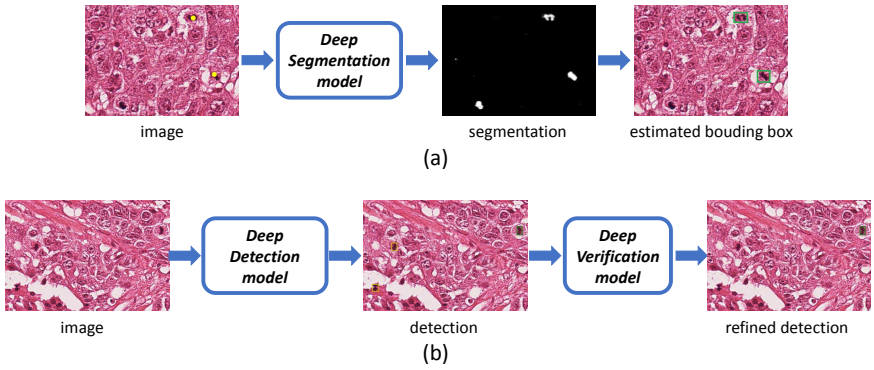

Figure 2: Mitosis detection system overview. (a) In training phase, it shows the process of generating estimated bounding box label for MITOSIS 2014 dataset. The yellow dot denotes the original centroid annotation, while the green box denotes the estimated bounding box using the segmentation result of the deep segmentation model. (b) In testing phase, it shows mitosis detection pipeline which includes a deep detection model followed by a deep verification model. The yellow box and green box are false positive and true positive, respectively.

detector. It provides a filtering system to reject the false positives misclassified by deep detection model, because its training data includes the hard negatives of detection model. Thus our system gets the bootstrapping mechanism to learn from mistakes.

The rest of this paper is organized as follows. A brief review of related work on mitosis detection, deep detection, segmentation and verification methods is given in Section 2. The following Section 3 introduces the details of our proposed approach. Section 4 details the experiments and shows state-of-the-art results on two publicly available datasets, whereas conclusions are made in Section 5.

2. Related work

There have been many approaches proposed for automatic mitosis detection from images. In terms of the image features, we can divide them into two types, handcrafted features based and CNN features based. Handcrafted features are widely applied in this problem to describe the appearance of mitotic cells (Veta et al., 2013; Khan et al., 2012; Wang et al., 2014; Sommer et al., 2012; Huang

& Lee, 2012; Malon et al., 2013; Irshad et al., 2013; Tek et al., 2013; Paul & Mukherjee, 2015). The handcrafted features usually contain shape, statistical and textural features to describe the mitotic cells. They are designed based on the domain knowledge of pathologists to recognize mitosis. The features are classified by support vector machines (SVM), random forest etc. The drawback of the handcrafted features is that they can not well describe the appearance of mitosis. Since there are a variety of morphologies and textures in mitotic cells, it is hard to manually design features to describe all mitosis very well. Another type of features used in this task is based on CNNs (Malon et al., 2013; Wang et al., 2014; Cireşan et al., 2013; Chen et al., 2016a). Compared with the handcrafted features, convolutional features are more powerful since they learn the representation of mitosis automatically. The disadvantage of deep convolutional features is its complexity in computation and a relatively long time to train. (Malon et al., 2013) combines manually designed features with CNN features and yields a 0.659 F-score on 2012 MITOSIS dataset. Most of the mitosis detection methods firstly generate candidates and then classify them by various classifiers to single out mitotic cells. Unlike this general scheme, (Cireşan et al., 2013) does not resort to the candidate segmentation process and directly applies the deep network classifier to images. It yields the highest F-score at 2012 MITOSIS contest and AMIDA13 challenge among all methods of participants. However, this pixel-wise classifier uses a sliding window way which is very computationally intensive in the test stage. Hence it is not very practical in clinic. Wang et al. (Wang et al., 2014) design a cascade system that requires fewer computing resources compared with (Cireşan et al., 2013). It leverages both handcrafted features and CNN features, and each type of features is used to train an individual classifier. In the test stage, images are classified by the two classifiers individually, and once the detection results of the two classifiers are not consistent, the image would be further classified by a second-stage classifier which is trained with both two features. However, it is not a pure deep learning based method. Selecting candidates is still based on a traditional cell segmentation method using Laplacian of Gaussian (LoG) responses on color

ratios, which is prone to losing mitosis since the handcrafted features could not accurately reflect the characteristics of mitosis. They only use the convolutional neural network to classify the candidate patch, and the network architecture is relatively small which results in the capacity of discrimination is not strong. So they use handcrafted features and traditional classifier to confirm the accuracy. CasNN (Chen et al., 2016a) leverages two convolutional neural networks to make up a deep cascaded detection system: a coarse retrieval model locates candidates through fully convolutional networks, and then a classification network is applied to find out mitosis from the candidates. Though the CasNN (Chen et al., 2016a) produces candidates by a retrieval neural network, its two neural networks (retrieval model and discrimination model) are trained independently. It is not trained in an end-to-end way, which impedes the integration of the two networks. Different from previous methods, we use the deep detection model to produce proposals and classify them in a single network, and the region proposal network and the subsequent classification network in deep detection model share full-image convolutional features. It means that the sections of candidates generation and classification can be trained jointly in an end-to-end fashion.

Deep learning based methods have significantly improved the accuracy of object detection and image classification (Krizhevsky et al., 2012). The Region based convolutional neural network (R-CNN) (Girshick et al., 2014) uses region proposals produced by the selective search algorithm (Uijlings et al., 2013), and recognize the proposals by SVM with deep convolutional features. It is very slow because the R-CNN performs a ConvNet forward pass for each proposal. For accelerating the detection speed, Fast R-CNN (Girshick, 2015) computes features for an entire image and extracts the features of a proposal by a region of interest (RoI) pooling layer. But the proposal is still generated by a translational and external method, and it accounts for a high portion of processing time. To address this problem, Faster R-CNN applies a RPN to generate proposals and the RPN shares the convolutional features with the Fast R-CNN classification network. This detector achieves a frame rate of 5 fps. The R-CNN based methods have been applied in many detection tasks, such as pedestrian detection

(Zhang et al., 2016; Li et al., 2015) and cell-phone usage detection (Hoang Ngan Le et al., 2016).

The fully convolutional network is proposed for semantic segmentation task, and it achieves state-of-the-art performance in several segmentation datasets, including PASCAL VOC and NYUDv2 (Silberman et al., 2012). Inspired by FCN, holistically-nested edge detection (HED)(Xie & Tu, 2015) is proposed. It leverages the deeply-supervised nets (DSN) (Lee et al., 2015) to perform multi-scale and multi-level feature learning. The multi-scale results provide a more accurate edge pixel localization and refine the edge segmentation prediction.

Though there have been some CNN-based methods for mitosis detection, most of them use the convolutional neural networks as the classifier or feature extractor, and no detection network, like R-CNN, has been applied to this task directly. Our method applies a deep detection model to this task and addresses the problems we met. Moreover, we use a multi-stage system to raise the accuracy, where the prediction results of detection component and verification component are combined through a fusion mechanism.

3. Methods

In this section, we describe the DeepMitosis method in detail. Our DeepMitosis method mainly consists of three components: A deep detection model (DeepDet) produces primary detection results. A deep verification model (DeepVer) verifies these detections and eliminates false positives to improve the accuracy. In addition, for the weak annotations that do not give out pixel-level labels, we utilize a deep segmentation model (DeepSeg) to segment the images and obtain estimated bounding boxes annotations.

3.1. Using Deep Detection network to detect mitosis

We illustrate the architecture of DeepDet in Fig. 3. Our detection model is based on Faster R-CNN. It utilizes a RPN to generate object location proposals which are category-agnostic. Over the last convolutional feature map, reference

boxes with different scales and aspect ratios are generated at each position. These reference boxes are called anchors. Two sibling fully-connected layers are then responsible for classifying the anchor and regressing the bounding box, respectively. Since these two fully-connected layers are not related to the spatial position, they are implemented as convolutional layers. For each anchor box, the anchor classifier predicts its probability of being a foreground object and the bounding box regressor outputs the estimated coordinates encoding the predicted bounding box of object. With the encoded coordinates, the anchor box can be transformed to a region proposal. Then, for each object proposal, a RoI pooling layer is applied to extract a fixed-length feature vector from the feature map. The RoI pooling leverages max pooling to transform the features inside the proposal to a fixed spatial extent feature map. Through the RoI pooling, the feature of any size proposal can be converted to a fixed size, which is required by the fully-connected layers of region classification network. Then the feature vector is fed into the region-based classification network which is used to recognize the proposals. The convolutional features of the classification network are shared with the RPN to reduce the computational cost. The classification network also has two sibling output layers, one for the probability estimates of each class, and another for predicting the encoded bounding boxes of each class.

The anchors of RPN have three scales and three aspect ratios. In our issue, the shape of mitosis is irregular, so it is necessary to apply multiple aspect ratios. We follow the default three aspect ratios of anchors in RPN, which are 1:1, 1:2 and 2:1.

Since the resolution of HPF images is very large, it is not convenient to utilize the original images to train the detector directly. We crop image patches from original images. Extracting sample is also a type of data augmentation since the cropped image patches are highly overlapped.

The mitosis in the original image is relatively small with an average side of 30 pixels, while the total stride in the last convolutional feature map is 16 pixels, so the region of a mitosis would be too coarse in the feature map, which is not suitable for fine-grained recognition. To address this problem, we re-scale the

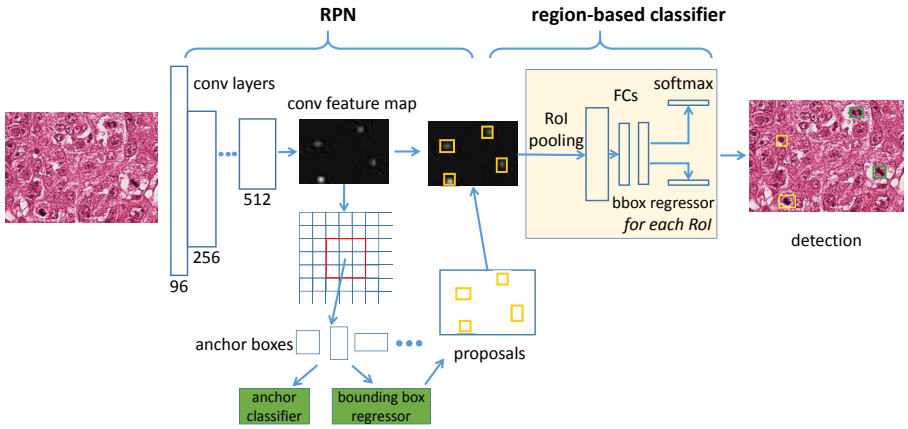

Figure 3: The architecture of DeepDet model. It consists of a RPN and a region-based classification model. It takes a histology image and generates convolutional feature maps. Upon the last feature map, anchors with different scales and aspect ratios are produced at each location. These anchors are handled by anchor classifiers and bounding box regressor to produce proposals. For each proposal, the RoI pooling layer extracts a fixed-length feature vector from the feature map. Then the region-based classifier predicts the score and output the refined bounding box position for each region proposal. The yellow box and green box in the last detection image are false positive and true positive, respectively.

image patch by two times. Thus the 16 pixels stride in the last convolutional feature map is equivalent to 8 pixels of the original image. The DeepDet uses 3 scales with box areas of 128^2, 256^2, and 512^2 pixels for detecting objects of different scales. An appropriate setting of scales is important for training the detection model. A statistic about mitotic cell area shows that there are few large mitotic cells, so we remove the largest scale 512 and add a small scale 64. The modified anchors can effectively cover mostly mitotic cells.

DeepDet is trained in an end-to-end way and the image mini-batch size is one. 256 anchors from an image are selected during a training batch. If an anchor is overlapped with any ground-truth bounding box and the Intersection-over-Union (IoU) is higher than 0.7, it would be assigned as a positive sample. On the contrary, if the IoU of an anchor is lower than 0.3 for all ground truth bounding boxes, the anchor would be assigned as a false sample. The anchors with IoU between 0.3 and 0.7 are ignored during training because they are not typical samples and prone to introduce confusion.

For each anchor, the RPN predicts its category (object or not object) and bounding box regression. Through the regression, anchors are converted to proposals for the down-stream region-based classification network. A proposal is labeled as foreground if its IoU with a ground-truth bounding box is not less than 0.5. And if its maximum IoU with any ground truth is in the interval $[0.1, 0.5)$, it would be labeled as background. In this training way, the chosen background RoIs are all overlapped with ground truth, while most background region would not contribute to the training. Considering that the number of mitosis in each image is low, and there are many difficult background regions should be taken into account in training. We change the IoU interval of background to $[0, 0.5)$, such that the negative proposals can be obtained from the background region by choosing the image patches that are not overlapped with mitosis.

3.2. Refining detection results by Deep Verification network

There may be many false positives in the detection results of DeepDet, so we take advantage of a verification network to classify the detections and eliminate

the false positives among them.

The DeepDet performs detection on the histology images and yields results. For the detector trained with estimated bounding boxes, its detections are not very reliable. Thus we develop a verification model following the DeepDet to refine the detection results. Another motivation to apply the verification model is that it can provide hard examples mining for the overall system. Faster R-CNN does not use bootstrapping and hence lacks the ability of mining hard examples (Shrivastava et al., 2016). During the training of DeepDet, the RPN proposals that are wrongly classified by the region-based classification model in one training iteration could not be collected for training the model in more iterations. Actually these hard samples are very valuable for the model optimization and can greatly enhance the discriminative capacity of model. Our verification model is trained on the detection results of DeepDet model including false positives, so it can obtain the capacity to identify hard mimics and thus can be viewed as a way of hard example mining.

We collect all the detections produced by DeepDet to train the DeepVer model. For a detected patch, we keep its centroid unchanged and extend the image patch to a square box with a fixed side length. Our DeepVer model is illustrated in Fig. 4. It is based on the ResNet (He et al., 2016), which achieves state-of-the-art performance in many vision tasks, such as ImageNet classification (Russakovsky et al., 2015b), ImageNet detection, and COCO detection (Lin et al., 2014). ResNet learns residual functions with reference to the layer inputs, instead of directly fitting a desired underlying mapping. The shortcut connection denotes an identity mapping, which sends the input to add with the output of stacked layers. The deep residual learning net can solve the degradation problem during training a substantial deep network and enjoys performance gains from increased depth.

The DeepVer model gives a probability score for each image patch. We fuse the scores of the DeepDet and the DeepVer. The final score of an image patch is

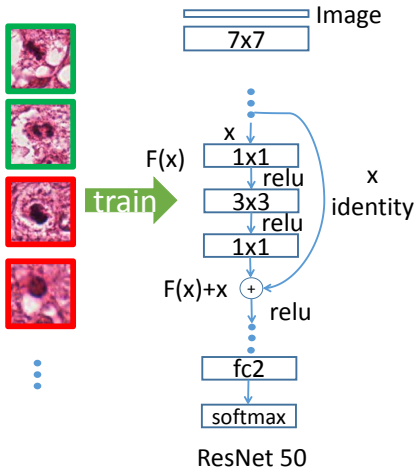

Figure 4: The architecture of DeepVer model. It is based on ResNet. The detection results of DeepDetare used for training the verification model. The image patches with green border and red border are positive samples and negative samples, respectively. Here we illustrate a building block of ResNet. The F(x) is the residual function that the stacked nonlinear layers need to fit. The shortcut connection simply performs identity mapping, and its output is added to the residual mapping F(x). A deep ResNet is constructed by stacking the blocks.

a weighted sum of the detector score $S_{DeepDet}$ and the classifier score $S_{DeepVer}$.

$$S = \omega \times S_{DeepDet} + (1 - \omega) \times S_{DeepVer} \tag{1}$$

This fusion method utilizes not only the classification score of DeepVer but also the prediction score of DeepDet when making decisions. It can take full advantage of predictions of the two models and explore the complementary of the two predictions to boost the accuracy.

3.3. Estimating bounding box label through Deep Segmentation network

In an object detection problem, the annotations are usually in bounding box format. The bounding boxes are necessary to train the DeepDet model. As shown in Fig. 5, there are two types of annotations in the mitosis datasets. One is the pixel-level ground truth that annotates every pixel of a mitotic cell. This type of label provides sufficient information that we can easily obtain an accurate bounding box for mitosis. Another kind of annotation merely labels the centroid of a mitosis and we can not get an accurate box to bound the mitosis. A simple strategy is to generate a fixed rectangle box for each mitosis. However, there is a wide range of aspect ratios and scales in mitotic cells. Thus the uniform square bounding box can not match the mitotic cells well. To solve this problem, we utilize a FCN segmentation model to process the images and then estimate the mitotic region based on the segmented images. This estimation provides pixel-level annotations hence we can obtain a refined bounding box label.

The FCN model is derived from VGG 16-layer net by replacing all fully connected layers with convolutional layers, and it produces pixel-wise prediction through a deconvolutional layer. We train a FCN segmentation model on the mitosis data that has pixel-level annotation, namely the 2012 MITOSIS dataset. For adapting the original FCN model (Long et al., 2015) to mitosis data, we modify the channel number of prediction layers to two, which represents mitosis and non-mitosis. Since the annotation has given the label of each pixel, we can easily transform them to label images, where 1 denotes the mitotic pixel and 0 represents a pixel of other cells. With the label images, we train a semantic

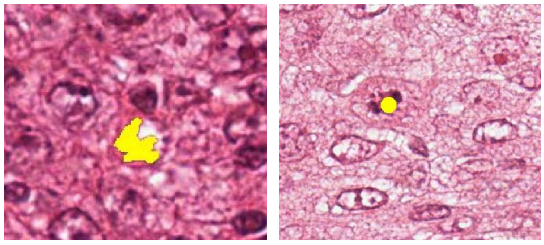

Figure 5: Annotations of mitosis datasets. The pixel highlighted with yellow is the annotations. (a) shows the pixel-level annotation of 2012 MITOSIS dataset which annotates every pixel of mitosis region. (b) shows the label of 2014 MITOSIS dataset that roughly gives the centroid of a mitotic cell. For better view, the single centroid pixel is enlarged to a circle.

segmentation model. After training such a model, we apply it to the 2014 MITOSIS dataset. The DeepSeg model predicts the regions of mitosis and gives refined bounding boxes of mitotic cells. The processing of producing bounding box of mitosis is illustrated in Fig. 6. DeepSeg model performs semantic segmentation on the histology image and predicts the mitotic regions. For each original centroid annotation, we locate a segmented mitotic blob covering the centroid pixel, and then use a rectangular box to bound the blob. This box is the refined bounding box label.

3.4. Mitosis detection on 2012 MITOSIS dataset and 2014 MITOSIS dataset

The annotations of the 2012 MITOSIS dataset and 2014 MITOSIS dataset are different, which results in the pipelines of proposed detection system on these two datasets are also different, as shown in Fig. 7. Specifically, the 2014 dataset needs two more steps: 1) it needs to be segmented by DeepSeg model for yielding bounding box annotations; 2) the detection results of DeepDet model need to be verified by the DeepVer model for further refinement. Though we perform semantic segmentation on 2014 MITOSIS dataset by DeepSeg model, the segmentation results maybe not very accurate, and that would result in inferior estimated bounding box annotations. The impure annotations damage the detection network optimization and limit the accuracy of detection. For

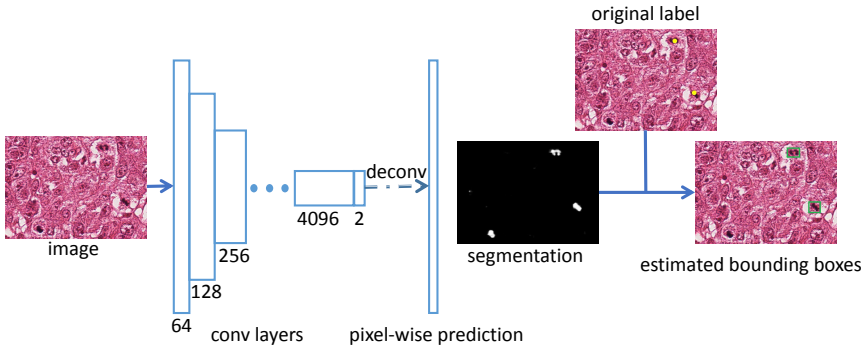

Figure 6: The processing of transforming centroid annotations to bounding box annotations by DeepSeg model. DeepSeg can produce pixel-wise prediction through a deconvolutional layer. For each mitosis in ground truth, we locate the corresponding segmented mitotic region and generate a bounding box from it. The green box denotes the estimated bounding box.

refining the results, we add the verification model following the detection network. While in 2012 dataset, the given pixel-level label has provided accurate and reliable bounding box, so we can obtain a powerful enough detector and do not need to deploy verification model.

4. Experiments and Results

In this section, we evaluate the performance of our proposed method for mitosis detection on 2012 MITOSIS contest dataset and 2014 MITOSIS dataset. On the 2012 MITOSIS dataset, we only exploit the DeepDet model since the pixel-level annotations are given. For the 2014 dataset, we use the overall system including segmentation model, detection model and verification model. The whole DeepMitosis system is implemented based on the Caffe deep learning framework (Jia et al., 2014) using Python and C++. The source code will be relsearsed on publication. Experiments are carried out on a Linux server with one NVIDIA GeForce GTX TITAN X GPU.

4.1. Datasets

2012 ICPR MITOSIS Dataset . The 2012 ICPR MITOSIS dataset (Roux et al., 2013) has 50 histopathology images corresponding to 50 HPFs at 40X

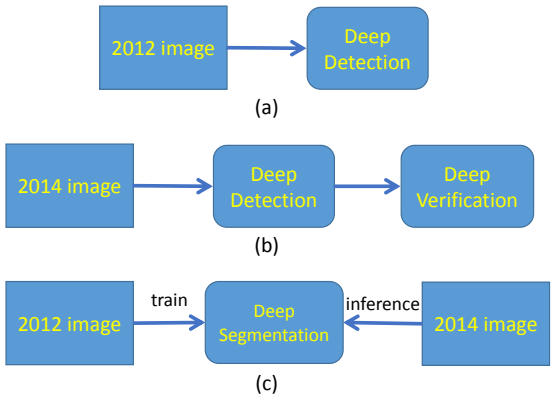

Figure 7: The pipelines of two mitosis dataset. (a) shows the detection pipeline of 2012 MITOSIS dataset, which merely utilizes the DeepDet model. (b) shows the detection pipeline of 2014 MITOSIS dataset. (c) shows that the DeepSeg model is trained on 2012 MITOSIS dataset and applied on 2014 dataset.

magnification, which are selected from breast cancer biopsy slides by pathologists. There are more than 300 mitotic cells labeled in the dataset and all pixels of each mitosis are annotated. We evaluate our method on the most widely used images produced by the Aperio XT scanner. The resolution of the scanner is 0.2456 μm per pixel. The HPF image has an area of 512×512 μm^2, so the size of each image is 2084×2084 pixels.

Following the rule of 2012 ICPR MITOSIS contest, 35 HPFs are used for training, and the remaining 15 HPFs are applied for testing. There are 226 and 101 mitosis in the training set and test set, respectively.

2014 ICPR MITOSIS Dataset . The 2014 ICPR MITOSIS dataset (MITOS-ATYPIA-14, 2014) has significantly more images than the 2012 MITOSIS dataset. It includes 1696 HPFs at 40X magnification. The size of each HPF is 1539×1376 pixels in this dataset. The training data consists of 1200 HPFs with 749 mitotic cells labeled, while in test set there are 496 HPFs but the number of mitotic cells is unknown since the annotation is held out by organizers. The mitotic cells in training data are labeled by two pathologists, and if there exists a conflict between them, another pathologist will give annotation and the final result is

determined by the majority. As mentioned above, the annotation in this dataset only labels the centroid rather than all pixels for each mitotic cell.

Performance measurements. The number of mitoses is critical in the breast cancer grading system, so the measurement of performance in the mitosis detection task is mainly based on the number of mitoses correctly detected, rather than the shape of detected mitosis. According to the contest criteria, the correct detection is the one that lies within a distance from the centroid of a ground truth mitosis. The distance is 5 μm (20 pixels) in 2012 MITOSIS contest and 8 μm (32 pixels) in 2014 MITOSIS contest.

Here we define some measures used for evaluating the accuracy of mitosis detection. D is the count of mitosis detected by our proposed approach. TP is the number of detections that are ground truth mitosis among the D mitosis detected, while FP is the number of detections that are not ground truth mitosis. And the number of ground truth mitosis not detected is defined as FN. With these measures, we can calculate the $recall$, $precision$, and $F-score$ using the following formulations:

$$recall = TP/(TP + FN) \tag{2}$$

$$precision = TP/(TP + FP) \tag{3}$$

$$F - score = 2 \times recall \times precision/(recall + precision) \tag{4}$$

4.2. Deep Detection model on 2012 MITOSIS dataset

Thanks to the precise and detailed ground truth of 2012 MITOSIS dataset, we can easily obtain the required bounding box annotations to train the DeepDet model. The model can yield excellent performance and we do not resort to verification model on this dataset.

Table 1: The performance of our detection models on 2012 MITOSIS test set. These models are trained using image patches with different scales. The "Patch Size" is the original size of cropped patch. All patches are uniformly re-scaled to 1024×1024 pixels to train DeepDet models, and the corresponding enlarged ratio of the image patch is shown in "Scale" row.

Patch Size	1024	640	512	256
Scale	1	1.6	2	4
F-score	0.568	0.762	**0.768**	0.756

4.2.1. Hyper-parameters

Our DeepDet model is based on VGG_CNN_M_1024 model (Chatfield et al., 2014), which is pre-trained on ImageNet classification dataset (Russakovsky et al., 2015b). We first train the model with the initial learning rate of 0.001 for 50k iterations, then continue training for 50k iterations with the learning rate of 0.0001, finally training for 20k iterations with the learning rate of 0.00001. We set momentum to 0.9, set weight decay to 0.0005, and batch size to 1.

4.2.2. Data augmentation of training data

Even the TITAN X GPU has 12GB memory, the full sized HPF image is too large to be taken as the input of the DeepDet model. Thus we need crop image patches from the original HPF images. In addition, a mitotic cell is relatively small under the original image scale, so we need to enlarge the images to fit the detection model which is initially designed for general object detection. For seeking the appropriate scale, we crop patches of different sizes from the original histology images and re-scale them to 1024×1024 pixels uniformly. Then we use these different scales patches to train the DeepDet model individually. Table 1 shows the performance of our DeepDet models trained on image patches with different scales.

It can be observed that the F-score is very low when using image patches of the original scale. And the performance improves drastically as the scale increases to 1.6. Further increasing the scale does not bring a distinct effect

Table 2: The performance of DeepDet models trained with different training sets. The performance is evaluated on 2012 MITOSIS test set.

Training set	F-score
AreaTh(800)	0.768
AreaTh(1600)+Rot	0.825
AreaTh(1000)+Rot	0.814
AreaTh(800)+Rot	**0.832**

on the F-score. For simplicity, we choose the scale 2 in our experiments as it achieves the best result among different scales. Moreover, using the moderate scale is relatively economical in memory. Specifically, we densely sample patches of 512×512 pixels from the original images with a step size of 32 pixels. Then, we re-scale the sampled image patches to 1024×1024 pixels. The mitotic cells in the boundary of patches may be split into two or more small parts. If the generated incomplete mitosis part is tiny, it should not be regarded as a valid image patch. We remove the image patches that contain small cross-boundary mitotic cells with an area lower than 800 pixels. The moderate threshold keeps effective positive samples as much as possible, and simultaneously filters out tiny mitosis parts which are prone to introduce error terms in training.

Training deep CNNs needs a large number of samples, so we rotate and mirror the original HPF images and then extract patches from the transformed images to produce more training samples. Here we rotate the original images in a step of 45 degrees. The data augmentation can yield more mitotic samples, which are critical for the training of the DeepDet network, especially when the number of mitotic samples in the original training dataset is significantly small.

Table 2 shows the performance of DeepDet models trained on different training sets. The number in "Training set" row is the area threshold used when filtering the boundary mitosis. For instance, the "AreaTh(800)" training set removes the image patches containing a boundary mitotic region smaller than 800 pixels. The "+Rot" means the training set includes rotated images;

otherwise, its data augmentation only includes image mirror. We note that the image rotation improves the F-score remarkably. Fig 8 shows detection results of "AreaTh(800)+Rot" detector on two HPFs of 2012 MITOSIS test set.

4.2.3. Parameters studies

We carry out some controlled experiments to examine how each parameter affects the performance of DeepDet model.

Anchor batch size . In the RPN training, every mini-batch contains positive and negative anchors from a single image. The "Anchor batch size" denotes the mini-batch size of anchors used in training. For the mitosis dataset, the number of positive samples in a single HPF image is often low. If the batch size is too large, the negative samples will dominate the data and bias the training. The default batch size is 256, and we evaluate some smaller batch size, e.g. 128 and 32. As reported in Table 3, the results of using small batch size are comparable to the best results, which indicates that the batch size is not very critical for RPN training and a relatively small value is also appropriate for this task.

Anchor scales. There are three anchor scales in the model, which are 128, 256 and 512. Considering that the mitotic cells are usually small even though the images are resized with scale 2, the largest anchor scale 512 is not appropriate for detecting mitosis. As shown in Table 3, the F-score improves 2% when we remove the scale 512 and add a smaller scale 64. It indicates that the anchor scale 64 is beneficial to the detection. Using merely two scales 64 and 128 are afford to give a good result. It makes sense since most of mitotic cells are in the range of the two scales.

Proposal number. The proposals generated by the RPN may overlap with each other, and the non-maximum suppression (NMS) is applied to reduce the number of proposals. In the detection stage, after the NMS, the top-N ranked proposals are selected based on the confidence generated by RPN. These proposals are then judged by the following classification sub-network. The parameter

Table 3: The performance of DeepDet models when training with different parameter configurations. All the detectors are trained on the optimal training set "AreaTh(800)+Rot" and the F-score is evaluated on 2012 MITOSIS test set.

Anchor Batch Size	Anchor Scales	Proposal Number	F-score
32	64,128,256	300	0.812
128	64,128,256	300	0.826
256	128,256,512	300	0.810
256	64,128	300	0.808
256	64,128,256	50	0.827
256	64,128,256	100	0.831
256	64,128,256	300	**0.832**

top-N is the "Proposal Number" in Table 3. We evaluate different numbers of proposals in experiments. Due to the low density of mitotic cells, the number of mitosis in a HPF image is usually small. Experimental results indicate that the performance is very robust to the number of proposal. Even we only use the top 50 proposals after NMS, it still achieves a 0.827 F-score, which is a state-of-the-art result as well.

4.2.4. Applying RPN to mitosis detection

Note that there is only one category of the foreground object in our issue, so we can apply a RPN to detect mitosis directly. As described above, the RPN has two sibling convolutional layers for classifying anchors and regressing bounding box. For our issue, the region proposals generated by the RPN can be viewed as the final detection results. The proposal scores can be taken as the prediction scores, while the regression predictions of proposals are the final bounding boxes predictions.

We choose the DeepDet model with the highest 0.832 F-score, and by removing its classification sub-network we can get the RPN model. As the parameters of fully connected layers in classification sub-network account for a majority of

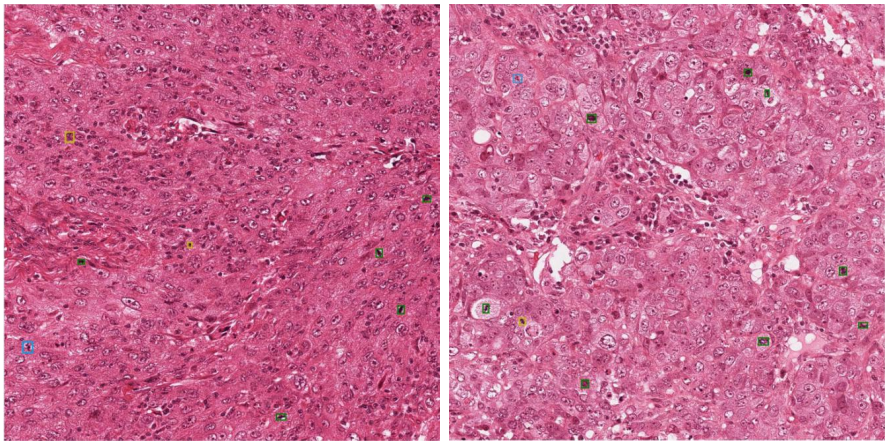

Figure 8: Mitosis detection results of DeepDet on two HPFs of the 2012 MITOSIS test set. Yellow, blue and green boxes denote false positives, false negatives and true positives, respectively.

proportion in the DeepDet, the size of the extracted RPN module is very small (about 30M). The F-score of this RPN model on the 2012 MITOSIS test set is 0.796. The competitive performance indicates that the extracted RPN model can be a good detector for this single-category detection problem. However, when we train a new RPN with our mitosis dataset independently, the F-score drops to 0.768. The nearly 3% gap indicates that the training mode of sharing convolutional features between the RPN and the classification subnetwork in DeepDet actually produces better convolutional features and higher performance.

4.2.5. Comparison with other methods

We then compare our method with some other approaches in performance. The details are shown in Table 4 and Fig. 9. Our approach achieves the highest F-score on 2012 MITOSIS test set. The IDSIA (Cireşan et al., 2013), IPAL (Irshad et al., 2013), SUTECH (Tashk et al., 2013), NEC (Malon et al., 2013) are the four best results attending the Mitosis detection contest in ICPR

Table 4: Performance of DeepDet with other competing approaches for 2012 MITOSIS test set.

Method	Precision	Recall	F-score
DeepDet	0.854	0.812	**0.832**
RRF (Paul et al., 2015)	0.835	0.811	0.823
CasNN (Chen et al., 2016a)	0.804	0.772	0.788
HC+CNN (Wang et al., 2014)	0.84	0.65	0.735
IDSIA (Cireşan et al., 2013)	0.886	0.70	0.782
IPAL (Irshad et al., 2013)	0.698	0.74	0.718
SUTECH (Tashk et al., 2013)	0.70	0.72	0.709
NEC (Malon et al., 2013)	0.75	0.59	0.659

505 2012. Among the CNN based methods (DeepDet, CasNN (Chen et al., 2016a), HC+CNN (Wang et al., 2014), IDSIA (Cireşan et al., 2013), NEC (Malon et al., 2013)), HC+CNN and NEC combine convolutional features and handcrafted features, and other CNN based methods utilize convolutional features only.

4.3. Mitosis detection on 2014 MITOSIS dataset

510 We use our DeepSeg model to segment histology images of 2014 MITOSIS dataset, and based on the segmentation results we estimate annotations of bounding box format. Then we train the DeepDet with these predicted annotations. Finally, to refine detections and filter out more false positives, we take advantage of DeepVer model to verify the results of DeepDet further.

515 We randomly sample 240 HPF images from the 2014 MITOSIS training data as the validation set, and the remaining 960 HPF images as the training set. There are 610 and 139 mitotic cells in the training set and the validation set, respectively.

4.3.1. Implementation of DeepSeg model

520 As described above, the annotation of 2014 MITOSIS dataset only labels the centroid pixel of a mitotic cell, and such annotation format can not train

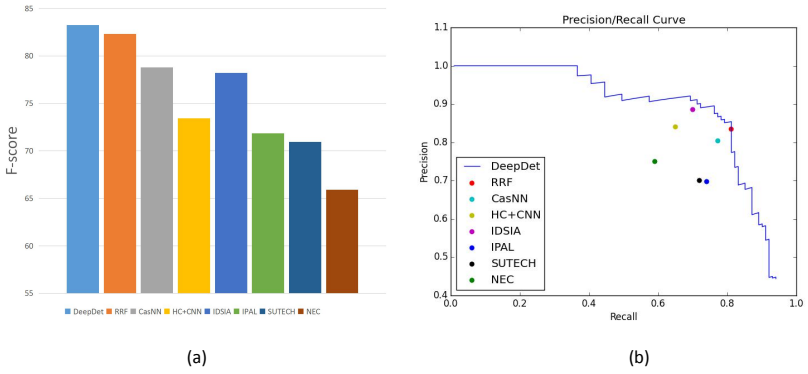

Figure 9: (a): F scores of DeepDet with other methods on 2012 MITOSIS test set. (b): Performance comparison with others in the PR plane.

an object detector directly. For achieving a finer bounding box ground truth, we leverage the DeepSeg model to segment the images and estimate a bounding box for each mitosis.

We use the 2012 MITOSIS dataset to train the DeepSeg model as it has pixel-level annotations. We sample patches of 521×521 pixels from HPF images. We augment the training data through mirroring and rotating the image patches. Since there are much more negative pixels than positive pixels, we perform more augmentation on patches containing positive pixels than the patches only having negative pixels, aiming at balancing the data. Binary label images can be easily generated based on the original pixel-level annotations to train the deep segmentation model.

The DeepSeg model is trained with Caffe framework (Jia et al., 2014). The FCN base model is trained with PASCAL semantic segmentation dataset. It is publicly available on model zoo site of Caffe. [1] We follow the default training configuration of FCN with learning rate 1e−10, momentum 0.99, weight decay 0.0005, and batch size 1. We also modify the output channel number of prediction layer to 2 for adapting the model to predict object or background.

[1] https://github.com/BVLC/caffe/wiki/Model-Zoo

According to the stride of prediction layer, the FCN models have three versions: FCN-32s, FCN-16s and FCN-8s. The number in the model name is the pixel stride at the final prediction layer, and the three models are generated in sequence (more details in (Long et al., 2015)). We first train the coarse one FCN-32s, then train the finer stride versions of FCN models. FCN-16s model is initialized with the parameters of the FCN-32s model we have trained. We change the learning rate to 1e−13 when training the FCN-16s model. Finally, we train the FCN-8s model based on the generated FCN-16s model with a learning rate 1e−14. We select the final FCN-8s model as the DeepSeg model.

The DeepSeg model is applied on the 2014 MITOSIS data and produces segmentation results. Since the spatial resolution of a HPF image is too big, we cut the image to 16 patches evenly when testing, and then stitch segmentation outputs of these patches to produce a full response image. We perform an adaptive binarization processing on the segmentation response map. The threshold is produced by Otsu's method (Otsu, 1975). For each mitosis centroid label, we use the segmented mitosis blob covering the centroid as the new annotation. We then get bounding box labels from the refined pixel-level ground truth. If there is no segmented mitotic region covering the centroid, we will assign a fixed bounding box to it.

4.3.2. Implementation of DeepVer model

Although we have refined the bounding box of mitosis, it is still not very accurate and reliable, which may introduce wrong supervision during the training. Here we exploit the DeepVer model to further verify the detection results of DeepDet. The verification model is based on the ResNet pre-trained on the ImageNet dataset (Russakovsky et al., 2015b). It is trained on the extracted detection patches of DeepDet. We keep the center of detection patch unchanged and extend the patch to 96×96 pixels. If the patch is positive, we will rotate the patch with a 90-degree step to produce more positive samples since the number of positive patches is relatively low. Meanwhile, we crop patches for each ground truth centroid, and perform the same rotation augmentation on them. More-

over, the original annotation has labeled some hard mimics which are prone to be wrongly identified as mitosis. We add these negative samples to our training data to improve the model capacity of recognizing the hard negatives. Finally, there are totally 16,248 image patches in the training set. We train a 50-layer ResNet as the DeepVer model. It can be seen as hard examples mining because the training samples are from the detection results of DeepDet. The learning rate is 0.01 and momentum is 0.9, and weight decay is 0.0001. We train the model for 12,000 iterations with a batch size of 64.

4.3.3. The effectiveness of DeepSeg and DeepVer models

We conduct some ablation experiments to demonstrate the effectiveness of our methods.

DeepDet (fixed). We first train a base DeepDet model on the 2014 MITOSIS-training set. Its parameters setting follows the configuration in 2012 MITOSIS dataset. We extract patches of 512×512 pixels from the training images and re-scale them to 1024×1024 pixels to train the DeepDet model. The anchor scales we used are 64, 128 and 256. We utilize a simple bounding box as the ground truth of mitotic cell. The side length of the box is fixed to 30 pixels. The originally labeled centroid pixel lies in the center of the box. Through careful validation on the dataset, the 30×30 square is found to be the best fixed annotation. We name the detector trained with fixed annotations as "DeepDet (fixed)".

DeepDet+Seg. This DeepDet model is trained with the same data and parameter setting as the DeepDet (fixed), but it leverages the refined bounding box annotations provided by the DeepSeg model. We name this model as "DeepDet+Seg". Table 5 shows that the DeepDet+Seg actually produces superior detection quality on 2014 MITOSIS validation set, compared with the DeepDet (fixed). This experimental result confirms the validity of refined annotations produced by the DeepSeg model.

DeepDet+Seg+Ver. we apply the DeepVer model to the detection patches produced by DeepDet+Seg in detection stage. We deploy a fusion method to combine the confidences of the detection model and verification model, rather than make judgments only by the latter one. This fusion strategy can exploit the valuable prediction scores given by DeepDet model. In our experiment, the fusion takes a weighted sum of the two model scores, and the weight is optimised on the validation set. We name this method as "DeepDet+Seg+Ver", as it utilizes the three deep models. It achieves a 0.582 F-score on the validation set as shown in Table 5. The nearly 8% improvement in performance compared to the DeepDet+Seg mainly results from the capacity of DeepVer model to distinguish negatives from detections. It demonstrates that the verification can filter out false positives of DeepDet effectively.

Table 5 shows that the performance increases, as more models be applied. With the segmentation model, the bounding box annotation can be finer and more accurate, and hence results in a better detector. And with the verification model, the false positives can be massively discarded, hence the accuracy advances remarkably.

Fig. 10 illustrates some detections examples of DeepDet+Seg and DeepDet+Seg+Ver in validation set. The first five columns confirm that the DeepDet+Seg+Ver can effectively identify false positives of DeepDet, due to the capacity of DeepVer model to distinguish negatives from detections. The sixth example shows a positive sample wrongly filtered by the DeepDet+Seg+Ver. The last column illustrates a mitosis missed by DeepDet+Seg is identified by the DeepDet+Seg+Ver. Fig. 11 shows the detection results of proposed DeepDet+Seg+Ver approach on two HPFs from the 2014 MITOSIS validation set.

4.3.4. Results on the test set

We then focus on the performance of our approach on the test set of 2014 MITOSIS dataset. The test set has 496 HPFs and no ground truth is given. Here we train the verification model using the detection patches of DeepDet+Seg from the training set and validation set. Experimental results on test set are

Table 5: Performance results of our methods on 2014 MITOSIS validation set.

Method	F-score
DeepDet (fixed)	0.489
DeepDet+Seg	0.505
DeepDet+Seg+Ver	**0.582**

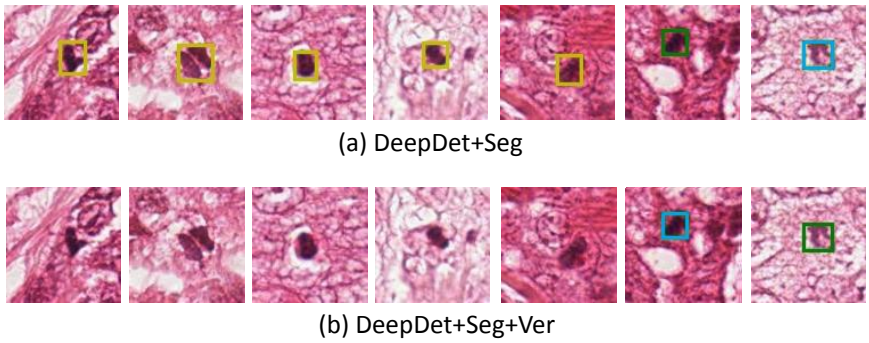

Figure 10: Example detections of our proposed methods on the validation set of 2014 MITOSIS dataset. (a) shows the detection results of DeepDet+Seg. (b) shows the results of DeepDet+Seg+Ver, which combines the predictions of detection model and verification model. We show seven examples to compare the differences between the two results. Yellow, blue and green boxes denote false positives, false negatives and true positives, respectively.

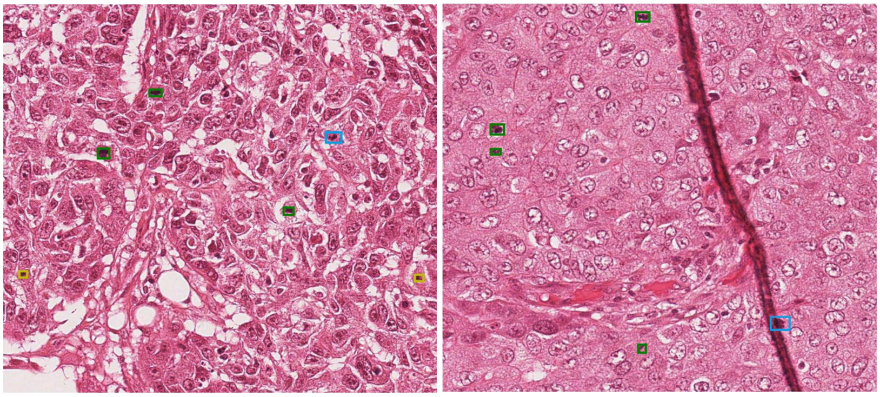

Figure 11: Mitosis detection results of DeepDet+Seg+Ver on two HPFs of the 2014 MITOSIS validation set. Yellow, blue and green boxes denote false positives, false negatives and true positives, respectively.

shown in Table 6. The DeepDet+Seg achieves a 0.398 F-score, which is a comparable result.

As expected, the DeepDet+Seg+Ver method achieves a higher F-score 0.437. Compared with the DeepDet+Seg, DeepDet+Seg+Ver has an identical precision, but a significant improvement on recall, which raises the F-score remarkably. The reason why the DeepDet+Seg has a lower recall is that it uses a relatively high detection score 0.984 as the threshold. The high threshold is a trade-off between the precision and recall. For example, if we take the score 0.9 as the decision threshold in DeepDet+Seg, its recall will increase to 0.522, but its precision will drop to 0.192, resulting in a much worse F-score 0.281. While for the DeepDet+Seg+Ver, we take the detections of Deep+Seg with score higher than 0.88 for the following verification of DeepVer. It can keep more positive detections and achieve a higher recall.

4.3.5. Comparison with other methods

The performance comparison of our proposed method with other approaches is reported in Table 7. Our method achieves a state-of-the-art performance with

Table 6: Performance results of our methods on 2014 MITOSIS test set.

Method	Precision	Recall	F-score
DeepDet+Seg	0.431	0.370	0.398
DeepDet+Seg+Ver	0.431	0.443	**0.437**

Table 7: Performance comparison of different approaches on 2014 MITOSIS test set. The first four methods are participants of 2014 ICPR MITOS-ATYPIA contest (MITOS-ATYPIA-14, 2014). "-" denotes the results which are not released.

Method	Precision	Recall	F-score
STRASBOURG	-	-	0.024
YILDIZ	-	-	0.167
MINES-CURIE-INSERM	-	-	0.235
CUHK	0.448	0.300	0.356
CasNN (Chen et al., 2016a)	0.411	0.478	0.442
DeepMitosis (DeepDet+Seg+Ver)	0.431	0.443	0.437

F-score 0.437, outperforming most of the methods except for the CasNN. The precision of our DeepMitosis (DeepDet+Seg+Ver) method is higher than that of CasNN, while our recall is inferior to the CasNN, which results in a 0.5% loss in F-score.

4.4. Time analysis

The target of automatic mitosis detection is to help expert pathologists in clinical applications. Since the number of HPFs in a single whole slide may be huge, it is significant to detect mitosis as quickly as possible. In our method, merely using the DeepDet model is able to obtain an excellent performance in 2012 MITOSIS dataset. Moreover, even the RPN module of our DeepDet model can produce a very good result on the 2012 MITOSIS dataset as well. For a HPF which has a spatial dimension 2084×2084 pixels, the DeepDet takes 0.72 s to detect mitosis and the RPN takes 0.68 s. The GPU we used in our experiment

is NVIDIA GeForce GTX TITAN X. As for the 2014 MITOSIS dataset, the system consists of two components in detection stage: detection model and verification model. The DeepDet+Seg takes 0.36 s per HPF in the test. The faster speed of detection model on 2014 image than 2012 Image results from the former has a relatively smaller resolution(1539×1376 pixels). The elapsed time of DeepVer depends on the processing speed on an image patch and the number of candidate patches DeepDet+Seg produced in a HPF. The DeepVer model takes 0.023 seconds per patch. On average, the DeepDet+Seg produces two detections with score higher than 0.8 per HPF. Hence the verification model takes about 0.05 seconds per HPF, and the total time of DeepDet+Seg+Ver is about 0.41 s for a HPF in 2014 MITOSIS dataset.

Compared with the IDSIA (Cireşan et al., 2013), which requires 31 s to apply a network on an input HPF image and 8 minutes to utilize two networks on eight variants for better performance, our approach outperforms it with a much faster speed. The efficiency of our method makes it more practical for clinical usage.

5. Conclusions and future works

In this paper, we propose a system named DeepMitosis for mitosis detection in H&E stained slide images. Our method leverages a deep detection model to perform detection. We adopt a general object detection method to the histopathology images and achieves excellent performance on 2012 MITOSIS dataset. It is noteworthy that merely applying a RPN can yield a comparative result. Since the 2014 MITOSIS dataset does not provide fine bounding box ground truth, we exploit a deep segmentation model to estimate the mitotic regions. The experimental results confirm that the segmentation result refines the annotations and improves the performance of detector. The effectiveness of the deep segmentation module indicates that it can significantly reduce the image labeling efforts in developing medical image analysis system based on deep learning. Meanwhile, we utilize a deep verification model to verify the results

of detection model, making up for the inferior capacity of the detector trained on the generated bounding boxes. The fused scores of the detector and the verification model can produce the state-of-the-art performance on the test set of 2014 MITOSIS dataset.

In future, we will explore methods that can produce more accurate pixel-wise labels for the centroid annotations, so that we can train more powerful detector on 2014 MITOSIS dataset to further improve the performance. Besides, we will study how to integrate the proposed DeepSeg, DeepDet and DeepVer networks into an end-to-end network trained using weak clinical annotations for accurate clinical diagnosis.

References

Chatfield, K., Simonyan, K., Vedaldi, A., & Zisserman, A. (2014). Return of the devil in the details: Delving deep into convolutional nets. *arXiv preprint arXiv:1405.3531*, .

Chen, H., Dou, Q., Wang, X., Qin, J., & Heng, P.-A. (2016a). Mitosis detection in breast cancer histology images via deep cascaded networks. In *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence* (pp. 1160–1166). AAAI Press.

Chen, H., Wang, X., & Heng, P. A. (2016b). Automated mitosis detection with deep regression networks. In *Biomedical Imaging (ISBI), 2016 IEEE 13th International Symposium on* (pp. 1204–1207). IEEE.

Ciresan, D., Giusti, A., Gambardella, L. M., & Schmidhuber, J. (2012). Deep neural networks segment neuronal membranes in electron microscopy images. In *Advances in neural information processing systems* (pp. 2843–2851).

Cireşan, D. C., Giusti, A., Gambardella, L. M., & Schmidhuber, J. (2013). Mitosis detection in breast cancer histology images with deep neural networks. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2013* (pp. 411–418). Springer.

Elston, C. W., & Ellis, I. O. (1991). Pathological prognostic factors in breast cancer. i. the value of histological grade in breast cancer: experience from a large study with long-term follow-up. *Histopathology*, *19*, 403–410.

Everingham, M., Zisserman, A., Williams, C. K., Van Gool, L., Allan, M., Bishop, C. M., Chapelle, O., Dalal, N., Deselaers, T., Dorkó, G. et al. (2007). The pascal visual object classes challenge 2007 (voc2007) results, .

Girshick, R. (2015). Fast r-cnn. In *Proceedings of the IEEE International Conference on Computer Vision* (pp. 1440–1448).

Girshick, R., Donahue, J., Darrell, T., & Malik, J. (2014). Rich feature hierarchies for accurate object detection and semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 580–587).

He, K., Zhang, X., Ren, S., & Sun, J. (2014). Spatial pyramid pooling in deep convolutional networks for visual recognition. In *European Conference on Computer Vision* (pp. 346–361). Springer.

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 770–778).

Hoang Ngan Le, T., Zheng, Y., Zhu, C., Luu, K., & Savvides, M. (2016). Multiple scale faster-rcnn approach to driver's cell-phone usage and hands on steering wheel detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops* (pp. 46–53).

Huang, C.-H., & Lee, H.-K. (2012). Automated mitosis detection based on exclusive independent component analysis. In *Pattern Recognition (ICPR), 2012 21st International Conference on* (pp. 1856–1859). IEEE.

Irshad, H. et al. (2013). Automated mitosis detection in histopathology using morphological and multi-channel statistics features. *Journal of pathology informatics*, *4*, 10.

Jia, Y., Shelhamer, E., Donahue, J., Karayev, S., Long, J., Girshick, R., Guadarrama, S., & Darrell, T. (2014). Caffe: Convolutional architecture for fast feature embedding. *arXiv preprint arXiv:1408.5093*, .

Khan, A. M., El-Daly, H., & Rajpoot, N. M. (2012). A gamma-gaussian mixture model for detection of mitotic cells in breast cancer histopathology images. In *Pattern Recognition (ICPR), 2012 21st International Conference on* (pp. 149–152). IEEE.

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems* (pp. 1097–1105).

Lee, C.-Y., Xie, S., Gallagher, P. W., Zhang, Z., & Tu, Z. (2015). Deeply-supervised nets. In *AISTATS* (p. 5). volume 2.

Li, J., Liang, X., Shen, S., Xu, T., Feng, J., & Yan, S. (2015). Scale-aware fast r-cnn for pedestrian detection. *arXiv preprint arXiv:1510.08160*, .

Lin, T.-Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., & Zitnick, C. L. (2014). Microsoft coco: Common objects in context. In *European Conference on Computer Vision* (pp. 740–755). Springer.

Long, J., Shelhamer, E., & Darrell, T. (2015). Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 3431–3440).

Malon, C. D., Cosatto, E. et al. (2013). Classification of mitotic figures with convolutional neural networks and seeded blob features. *Journal of pathology informatics*, *4*, 9.

MITOS-ATYPIA-14 (2014). Mitos-atypia-14-dataset. `https://mitos-atypia-14.grand-challenge.org/dataset/`. [Online; accessed 17.03.03].

Ning, F., Delhomme, D., LeCun, Y., Piano, F., Bottou, L., & Barbano, P. E. (2005). Toward automatic phenotyping of developing embryos from videos. *Image Processing, IEEE Transactions on*, *14*, 1360–1371.

Otsu, N. (1975). A threshold selection method from gray-level histograms. *Automatica*, *11*, 23–27.

Paul, A., Dey, A., Mukherjee, D. P., Sivaswamy, J., & Tourani, V. (2015). Regenerative random forest with automatic feature selection to detect mitosis in histopathological breast cancer images. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015* (pp. 94–102). Springer.

Paul, A., & Mukherjee, D. P. (2015). Mitosis detection for invasive breast cancer grading in histopathological images. *Image Processing, IEEE Transactions on*, *24*, 4041–4054.

Ren, S., He, K., Girshick, R., & Sun, J. (2015). Faster r-cnn: Towards real-time object detection with region proposal networks. In *Advances in neural information processing systems* (pp. 91–99).

Roux, L., Racoceanui, D., Loménie, N., Kulikova, M., Irshad, H., Klossa, J., Capron, F., Genestie, C., Le Naour, G., & Gurcan, M. N. (2013). Mitosis detection in breast cancer histological images an icpr 2012 contest. *Journal of pathology informatics*, *4*.

Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., Berg, A. C., & Fei-Fei, L. (2015a). ImageNet Large Scale Visual Recognition Challenge. *International Journal of Computer Vision (IJCV)*, *115*, 211–252. doi:10.1007/s11263-015-0816-y.

Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M. et al. (2015b). Imagenet large scale visual recognition challenge. *International Journal of Computer Vision*, *115*, 211–252.

Shrivastava, A., Gupta, A., & Girshick, R. (2016). Training region-based object detectors with online hard example mining. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 761–769).

Silberman, N., Hoiem, D., Kohli, P., & Fergus, R. (2012). Indoor segmentation and support inference from rgbd images. In *Computer Vision–ECCV 2012* (pp. 746–760). Springer.

Sommer, C., Fiaschi, L., Hamprecht, F. A., & Gerlich, D. W. (2012). Learning-based mitotic cell detection in histopathological images. In *Pattern Recognition (ICPR), 2012 21st International Conference on* (pp. 2306–2309). IEEE.

Tashk, A., Helfroush, M. S., Danyali, H., & Akbarzadeh, M. (2013). An automatic mitosis detection method for breast cancer histopathology slide images based on objective and pixel-wise textural features classification. In *Information and Knowledge Technology (IKT), 2013 5th Conference on* (pp. 406–410). IEEE.

Tek, F. B. et al. (2013). Mitosis detection using generic features and an ensemble of cascade adaboosts. *Journal of pathology informatics*, *4*, 12.

Uijlings, J. R., Van De Sande, K. E., Gevers, T., & Smeulders, A. W. (2013). Selective search for object recognition. *International journal of computer vision*, *104*, 154–171.

Veta, M., van Diest, P., & Pluim, J. (2013). Detecting mitotic figures in breast cancer histopathology images. In *SPIE Medical Imaging* (pp. 867607–867607). International Society for Optics and Photonics.

Veta, M., Van Diest, P. J., Willems, S. M., Wang, H., Madabhushi, A., Cruz-Roa, A., Gonzalez, F., Larsen, A. B., Vestergaard, J. S., Dahl, A. B. et al. (2015). Assessment of algorithms for mitosis detection in breast cancer histopathology images. *Medical image analysis*, *20*, 237–248.

Wang, H., Cruz-Roa, A., Basavanhally, A., Gilmore, H., Shih, N., Feldman, M., Tomaszewski, J., Gonzalez, F., & Madabhushi, A. (2014). Cascaded ensemble

of convolutional neural networks and handcrafted features for mitosis detection. In *SPIE Medical Imaging* (pp. 90410B–90410B). International Society for Optics and Photonics.

Xie, S., & Tu, Z. (2015). Holistically-nested edge detection. In *Proceedings of IEEE International Conference on Computer Vision*.

Zhang, L., Lin, L., Liang, X., & He, K. (2016). Is faster r-cnn doing well for pedestrian detection? In *European Conference on Computer Vision* (pp. 443–457). Springer.