



DeepMitosis: Mitosis detection via deep detection, verification and segmentation networks

Chao Li^a, Xinggang Wang^{a,*}, Wenyu Liu^a, Longin Jan Latecki^b

^aSchool of Electronics Information and Communications, Huazhong University of Science and Technology, Wuhan, PR China

^bCIS Department, Temple University, Philadelphia, PA 19122, USA

ARTICLE INFO

Article history:

Received 30 April 2017

Revised 11 October 2017

Accepted 2 December 2017

Available online 31 January 2018

Keywords:

Mitosis detection

Faster R-CNN

Fully convolutional network

Breast cancer grading

ABSTRACT

Mitotic count is a critical predictor of tumor aggressiveness in the breast cancer diagnosis. Nowadays mitosis counting is mainly performed by pathologists manually, which is extremely arduous and time-consuming. In this paper, we propose an accurate method for detecting 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 pixel of mitosis is annotated), an elaborately designed deep detection network for localizing mitosis by using contextual region information, and a deep verification network for improving detection accuracy by removing false positives. We validate the proposed deep learning method on two widely used Mitosis Detection in Breast Cancer Histological Images (MITOSIS) datasets. Experimental results show that we can achieve the highest F-score on the MITOSIS dataset from ICPR 2012 grand challenge merely using the deep detection network. For the ICPR 2014 MITOSIS dataset that only provides the centroid location of mitosis, we employ the segmentation model to estimate the bounding box annotation for training the deep detection network. We also apply the verification model to eliminate some false positives produced from the detection model. By 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.

© 2018 Elsevier B.V. All rights reserved.

1. Introduction

According to the Nottingham Grading System (Elston and Ellis, 1991), there are three important morphological features on Hematoxylin and Eosin (H&E) stained slides for breast cancer grading. They are mitotic count, tubule formation, and nuclear pleomorphism. Among them, mitotic count is the most important biomarker. Pathologists usually search for mitosis on high-power fields (HPFs) manually. It is a time-consuming and tedious task 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, making it hard to reach a consensus on mitotic count among pathologists. Thus it is very essential to develop automatic detection methods, 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 the following challenges. First, the appearance of mitosis varies in a wide range, as shown in Fig. 1(a). Mitosis has four phases called prophase, metaphase, anaphase and telophase. The shapes and structures of cells in different phases are diverse. In the telophase stage, the nucleus of 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. In the third row of Fig. 1(a), we show some mitotic cells in telophase stage. Second, mitotic cells are considerably less than non-mitotic cells. The low probability of their occurrence makes the detection more difficult. Third, there are some other cells (like apoptotic cells, dense nuclei) that have similar appearances with mitosis, making it hard to screen them out. Fig. 1(b) illustrates some non-mitotic cells.

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), the 2014 ICPR MITOS-ATYPIA challenge (MITOS-ATYPIA-14, 2014), and the mitosis detection task in TUPAC16 contest at MICCAI 2016 (TUPAC16, 2016). Many of the existing methods use handcrafted

* Corresponding author.

E-mail addresses: chaoli@hust.edu.cn (C. Li), xgwang@hust.edu.cn (X. Wang), liuwuy@hust.edu.cn (W. Liu), latecki@temple.edu (L.J. Latecki).

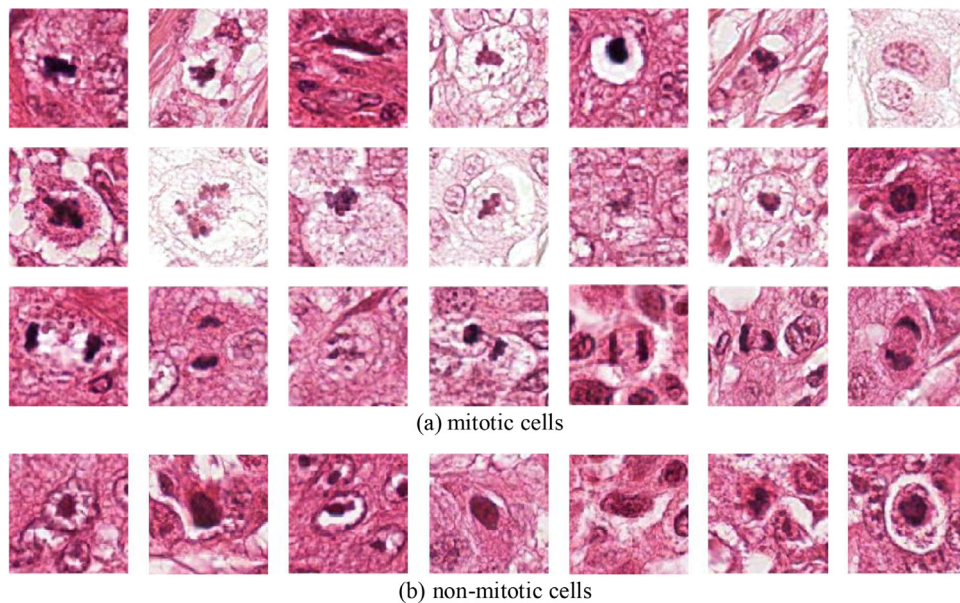


Fig. 1. Examples of mitotic cells and non-mitotic cells. (a) shows some mitotic cells. The third row contains mitoses in telophase stage. (b) shows some non-mitotic cells with similar appearances to mitotic cells.

features to model the appearance of mitotic cells. 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 explicitly capture characteristics of mitosis. However, due to the diverse and complex shapes of mitosis and the existence of confounding cells, it is hard to manually design discriminative features to effectively distinguish mitosis from non-mitotic cells.

Since the remarkable work of 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; Tang et al., 2017; Wang et al., 2018), object detection (Ren et al., 2015; Li et al., 2017) and semantic segmentation (Long et al., 2015). In biomedical analysis field, CNN based methods also yield excellent performance in several tasks, for instance, the brain tumor segmentation (Havaei et al., 2017), the analysis of developing embryos from videos (Ning et al., 2005), and so on. Hence some CNN based approaches (Malon et al., 2013; Cireşan et al., 2013; Wang et al., 2014; Chen et al., 2016a; 2016b) have been proposed to detect mitosis. Compared with traditional methods, deep learning based methods can learn more efficacious features from data automatically.

Current deep networks based mitosis detection methods typically fall into two categories. (1) Considering mitosis detection as a classification problem, e.g., IDSIA (Cireşan et al., 2013) 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, e.g. Chen et al. (2016b) infer the location of mitosis using fully convolutional networks (FCN), which ignores region information and makes it hard to deal with weak labels, e.g., the 2014 MITOSIS dataset that only labels the center of mitosis.

We argue that it makes more sense to consider the mitosis detection task as an object detection problem, and propose a method named DeepMitosis using deep detection network to solve this problem. Meanwhile, to the best of our knowledge, this is the first paper that applies a deep detection method to the mitosis detection problem.

Faster R-CNN (Ren et al., 2015), a region-based ConvNets detection model, uses a fully convolutional Region Proposal Network (RPN) to generate proposals and then applies a region-based classification network to classify these proposals. We adapt this deep detection model to the mitosis detection task.

An overview of the 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 (He et al., 2016) for classifying the detection patches to further improve the accuracy. The training of deep detector demands the bounding box labels. We can train the detector on 2012 MITOSIS dataset directly since this dataset has already given the label of each pixel. However, the 2014 MITOSIS dataset only annotates the centroid of mitosis, thus we need to estimate the bounding box annotations of mitotic cells before training the detector. Inspired by the success of FCN in semantic segmentation in natural images, we utilize a FCN model that has been trained on 2012 MITOSIS dataset to perform segmentation on the 2014 MITOSIS dataset. For each labeled centroid, we infer a bounding box by combining the segmentation result and the centroid label. The predicted box labels are then utilized to train the deep detector. The inference process has two stages, as shown in Fig. 2(b). 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 is a powerful classification network. The final result is a weighted sum of the predictions from the detection model and the verification model.

In summary, we have at least four main contributions in this paper. (1) We obtain the state-of-the-art results on two challenging mitosis detection datasets with high detection speed. (2) We tailor the general object detection framework Faster R-CNN to mitosis detection in H&E stained histopathological images. To our best knowledge, this work is the first time the proposal-based deep detection CNN is applied to the mitosis detection problem. (3) We propose a solution to train bounding box detector in a weakly supervised way, through using a deep segmentation

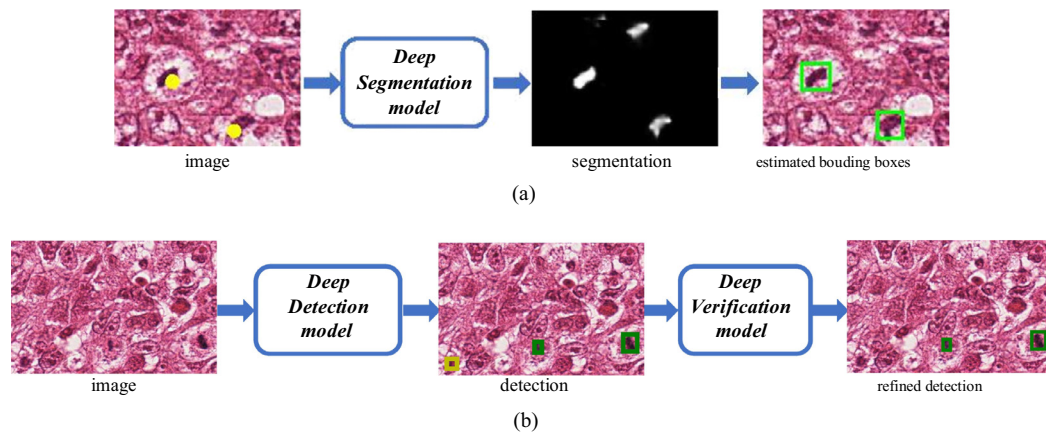


Fig. 2. Mitosis detection system overview. (a) In training phase, it shows the process of generating estimated bounding box label for 2014 MITOSIS dataset. The yellow dot denotes the original centroid annotation, while the green box denotes the estimated bounding box. The top segmented blob does not correspond to a marked centroid, so we discard it. (b) In testing phase, it shows the 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. A false positive is filtered out in the verification stage. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

network to estimate bounding box labels. This weakly supervised detector learning method helps to significantly reduce the labeling effort of pathologist. (4) We adopt a classification model to further verify the detection results. The deep verification model provides a bootstrapping mechanism for mining hard negative examples. Fusing the predictions of the deep detection model and deep verification model can improve the performance.

The rest of this paper is organized as follows. A brief review of related work is given in Section 2. Section 3 introduces the details of our proposed approach. Section 4 presents the experiments and shows state-of-the-art results on two publicly available datasets. Conclusions are made in Section 5.

2. Related work

There have been many approaches proposed for automatic mitosis detection. 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 and Lee, 2012; Malon et al., 2013; Irshad et al., 2013; Tek et al., 2013; Paul and Mukherjee, 2015). The handcrafted features usually contain shape, statistical and textural features. They are designed based on the domain knowledge of pathologists to recognize mitosis. The features are usually classified by support vector machines (SVM), random forest and Adaboost, etc. The drawback of the handcrafted features is that they fail to 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 accurately.

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 prove more powerful since they learn the representation of mitosis automatically. The disadvantage of deep convolutional features is their complexity in computation and a relatively long training time. Malon et al. (2013) combine manually designed features with CNN features and obtain a 0.659 F-score on 2012 MITOSIS dataset. IDSIA (Cireşan et al., 2013) directly applies a deep classification network to images in a sliding window way. It yields the highest F-score at 2012 MITOSIS contest and AMIDA13 challenge among all participating methods. However,

this detection way is very computationally intensive, which makes the method not practical in clinic. HC+CNN (Wang et al., 2014) designs a cascade system that requires fewer computing resources compared with IDSIA (Cireşan et al., 2013). It utilizes both handcrafted features and CNN features. In the test stage, images are classified by the handcrafted features based classifier and the CNN classifier, respectively. Once the detection results of the two classifiers are not consistent, the image would be further classified by a second-stage classifier trained with both features. 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. This candidate generation method is prone to missing mitosis since the handcrafted features cannot accurately describe the mitosis. The CNN they used is relatively small with weak capacity of discrimination. CasNN (Chen et al., 2016a) adopts two convolutional neural networks to make up a deep cascaded detection system: a coarse retrieval model to locate candidates through fully convolutional networks, and a classification network to find out mitosis from the candidates. Though the CasNN (Chen et al., 2016a) utilizes a neural network to produce candidates, its two components (retrieval model and classification 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 apply a single deep detection model to this problem. Our deep detection model is a pure deep learning based method, which can produce candidate proposals and classify them in a single model. Convolutional features are shared between the region proposal network and the subsequent classification network in this model. We train the two sub-networks 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 recognizes the proposals by SVM with deep convolutional features. R-CNN is very slow because it performs a ConvNet forward pass for each proposal. To accelerate the detection speed, Fast R-CNN (Girshick, 2015) computes features for the entire image and extracts the proposal features using a region of interest (RoI) pooling layer. However, the proposals are still generated by a traditional and external method, which accounts for a high portion of processing time. To address this problem, Faster R-CNN applies a RPN to generate proposals. 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. 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 and Tu, 2015) is proposed. It utilizes 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. Recently, some deep segmentation methods have been proposed for medical image analysis (Ronneberger et al., 2015; Roth et al., 2016; Christ et al., 2017; Havaei et al., 2017; Kamnitsas et al., 2017). U-Net (Ronneberger et al., 2015) extends the architecture of FCN by a symmetric expanding path. This network has shortcut connections from low layers (contracting path) to high layers (expanding path) that have equal resolution. Spatial information from low layers can be propagated to high layers so that the localization is more accurate. The work in Roth et al. (2016) utilizes HED (Xie and Tu, 2015) to learn organ interior and boundary mid-level cues for subsequent spatial aggregation in pancreas segmentation. The deeply learned mid-level cues are beneficial to the organ segmentation and lead to state-of-the-art results for pancreas segmentation in CT abdomen images. CFN (Christ et al., 2017) deploys cascaded FCNs for the combined automatic segmentation of the liver and its lesions. This method selects the U-Net architecture (Ronneberger et al., 2015) as the FCN segmentation model. A first U-Net segments the liver and then a second U-Net segments the lesions within the liver ROIs. Havaei et al. (2017) propose a two-pathway architecture for brain tumor segmentation. The two pathways respectively model local details and global context. However, this network predicts the class of a pixel with a patch centered at that pixel, which is less efficient.

Though there have been some CNN-based methods for mitosis detection, most of them use the CNN to classify patches or extract features. We observe that no detection network, like R-CNN, has been applied to this task. Our method adapts a deep detection model to this medical image task. Furthermore, we solve the problem of training the detection model on weakly annotated mitosis data by means of a segmentation model. In addition, we use a multi-stage system to improve the accuracy by fusing the predictions of detection component and verification component.

3. Methods

In this section, we describe the DeepMitosis method in detail. It 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. In addition, for the weak annotations that do not provide 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

DeepDet is the core component of our DeepMitosis system. We illustrate the architecture of DeepDet in Fig. 3. It utilizes a RPN to generate category-agnostic proposals. 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 responsible for clas-

sifying the anchor and regressing the bounding box, respectively. For each anchor box, the anchor classifier predicts its probability of being a foreground object, and the bounding box regressor estimates the location coordinates of the object. With the predicted coordinates, the anchor box can be transformed to a region proposal. Then, for each object proposal, a RoI pooling layer extracts a fixed-length feature vector from the feature map. The feature vector of the proposal is then fed into the region-based classification network. The classification network also has two sibling output layers, one softmax layer for estimating the classification probability, and a regressor layer for encoding the bounding boxes of each class. To reduce the computational cost, the convolutional features of the classification network are shared with the RPN.

The anchors of RPN have three scales and three aspect ratios. In this application, the shape of mitosis is irregular, making it 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 full image to train the detector directly. We extract highly overlapped patches from original images, which can be considered as a way of data augmentation.

Mitotic cell is relatively small with an average side of 30 pixels in HPF images. However, the stride in the last convolutional feature map of RPN is equivalent to 16 pixels in the original image. The coarse feature map is not suitable for accurate object localization, especially for the small mitoses. To address this problem, we up-sample the image patches by factor of two. Thus the 16 pixels stride in the last convolutional feature map is equivalent to 8 pixels in the original image. The Faster R-CNN uses 3 anchor scales with areas of 128^2 , 256^2 , and 512^2 pixels for detecting objects of different scales. Statistics about mitotic cell area show 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 most mitotic cells.

DeepDet is trained in an end-to-end way. 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.

RPN predicts the category (object or not object) and regresses the bounding box for each anchor. Through the regression, anchors are converted to proposals for the down-stream region-based classification network. To train the classification network, a proposal is labeled as foreground if its IoU with a ground-truth bounding box is not less than 0.5. If its maximum IoU with any ground truth is in the interval [0.1, 0.5), it is labeled as background. In this way, the chosen background proposals are all overlapped with ground truth, while most background region would not contribute to the training. Considering that the number of mitosis is small, many hard negatives in background region should be taken into account during the training. Therefore, we change the IoU interval to [0, 0.5). Thus the negative proposals can be selected from the background region not overlapping with mitosis.

3.2. Refining detection results by deep verification network

For the detector trained with estimated bounding boxes, its detection results may include many false positives. Thus we need a verification model following the DeepDet to refine the 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 lacks the capacity for mining hard

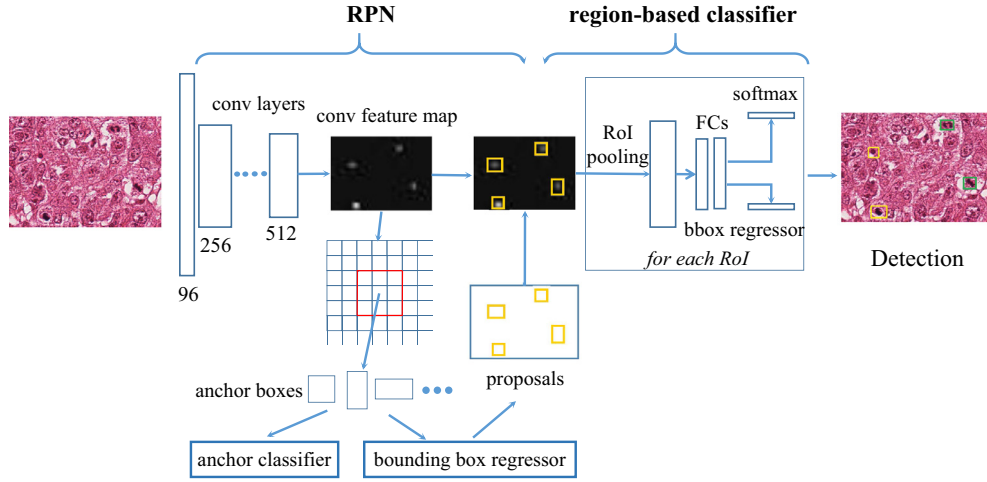


Fig. 3. The architecture of DeepDet model. It consists of a RPN and a region-based classifier. It takes a histology image and generates convolutional feature maps. Upon the last feature map, anchors with multiple scales and aspect ratios are produced at each location. These anchors are handled by anchor 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 outputs a score and a refined bounding box position for each region proposal. The green box and yellow box in the last detection image mark true positive and false positive, respectively. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

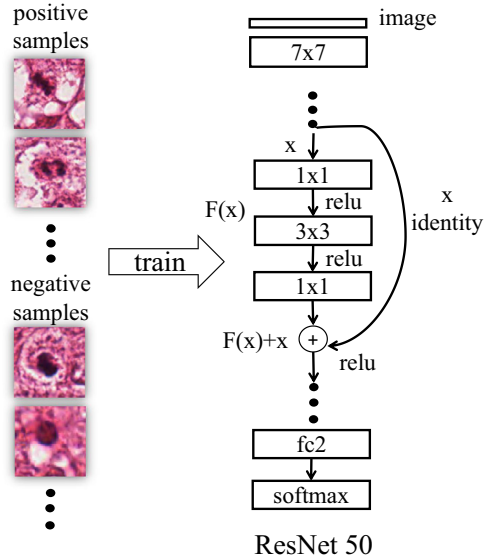


Fig. 4. The architecture of DeepVer model. It is based on ResNet. Here we illustrate a building block. The $F(x)$ is the residual function that the stacked nonlinear layers need to fit. The shortcut connection simply performs identity mapping, and its output is added to the residual mapping $F(x)$. A deep ResNet is constructed by stacking these blocks. DeepVer is trained with the detection results of DeepDet.

examples (Shrivastava et al., 2016). In the DeepDet training, the RPN proposals that are wrongly classified by the region-based classifier cannot be collected for the training of next iterations. In fact, these hard samples are very beneficial to the model optimization. They can greatly enhance the discriminative capacity of model. Our verification model is trained on the DeepDet's results, particularly the produced false positives, so it can learn the ability to identify the hard mimics. Thus the verification model can be viewed as a way of hard example mining for DeepDet.

We collect all the detections produced by DeepDet to train the DeepVer model. For each detection, we extract a square patch with fixed size. Our DeepVer model is illustrated in Fig. 4. It is based on the ResNet (He et al., 2016), which achieves state-of-the-art performances in many vision tasks, such as ImageNet classification (Russakovsky et al., 2015b), ImageNet detection,

and COCO detection (Lin et al., 2014). Instead of fitting a desired underlying mapping directly, the ResNet learns residual functions with reference to the layer inputs. The shortcut connection is an identity mapping, which sends the input to add with the output of stacked layers. Through this designed structure, deep residual learning net can solve the degradation problem when training a substantial deep network.

In the detection phase, the DeepVer model outputs a probability score for each detection patch of DeepDet. We design two strategies to utilize the DeepVer. The first method is a cascade way. It merely utilizes the classification scores of DeepVer to judge the detection patches. Another method fuses the detector score $S_{DeepDet}$ and the classifier score $S_{DeepVer}$, as shown below.

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

The weight ω is determined by a parameter sweep on a validation set. When making decisions, this fusion method utilizes not only the classification score of DeepVer but also the prediction score of DeepDet. As the DeepVer is optimized on the detection results produced by DeepDet, it may give low confidences for the false positives of DeepDet. Hence the verification phase can effectively eliminate false positives and improve the accuracy.

3.3. Estimating bounding box label through deep segmentation network

In the object detection problem, the annotations are usually in the format of bounding box. Training the DeepDet model also requires the bounding box label. As shown in Fig. 5, there are two types of annotations in mitosis datasets. One is the pixel-level ground truth that annotates every pixel of a mitotic cell. This type of label provides sufficient information to easily obtain an accurate bounding box for mitosis. Another kind of annotation merely labels the centroid pixel of a mitosis and we cannot get an accurate box to bound the mitosis tightly. A straight-forward method is to produce a fixed rectangle box for each mitosis. However, there is a wide range of aspect ratios and scales in mitotic cells. A fixed bounding box cannot enclose every mitotic cell very well. To solve this problem, we utilize a FCN model to segment the images and estimate the mitotic region. A refined bounding box label can be obtained from the predicted mitotic region.

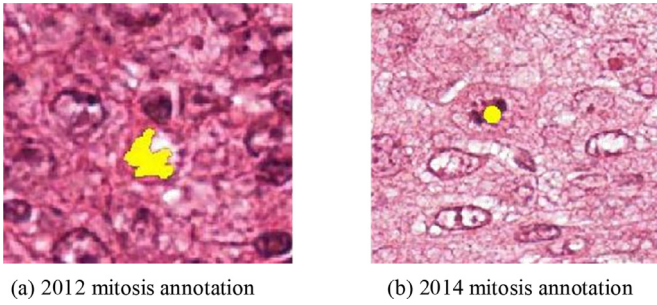


Fig. 5. Annotations of mitosis datasets. The pixels highlighted with yellow are the annotations. (a) shows the pixel-level annotation of 2012 MITOSIS dataset which annotates every pixel of mitosis region. (b) shows the label of 2014 MITOSIS dataset that roughly gives the centroid of a mitotic cell. For a better visualization, the single centroid pixel is enlarged to a circle. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

The FCN model is derived from VGG 16-layer net by replacing all fully connected layers with convolutional layers. It can produce pixel-wise prediction through the deconvolutional layer. To train the segmentation model, we use the pixel-level annotated images from the 2012 MITOSIS dataset. To adapt 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. We produce label images from the pixel-level annotations for training the semantic segmentation model. After training such a model, we apply it to the weakly annotated 2014 MITOSIS dataset to generate refined bounding boxes for mitotic cells, as illustrated in Fig. 6. DeepSeg model performs semantic segmentation on the histology image and estimates the mitotic regions. For each original centroid annotation, we locate a segmented mitotic blob that covers the centroid pixel, and then use a rectangular box to bound the blob. This box is the refined bounding box label. The segmented blobs that do not contain any centroid label are ignored. If an original centroid label does not have any segmented blob to cover, it will be assigned a fixed bounding box.

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 leads to two different pipelines of the proposed system, as shown in Fig. 7. Specifically, the 2014 dataset needs two more steps: (1) the segmentation step uses the DeepSeg model to generate bounding box annotations; (2) the verification step uses the DeepVer model to further verify the detection results.

The original annotation of 2014 MITOSIS dataset only contains the centroid pixel, so we resort to the DeepSeg to estimate the bounding box of mitotic cell. However, the estimated bounding box annotation is not accurate enough, which may limit the detection accuracy of the DeepDet model. So we add the verification model following the detection network to improve the detection result. While in 2012 dataset, its pixel-level label brings about accurate and reliable bounding box, so we can obtain a powerful enough detector and produce outstanding detection results without any verification model.

4. Experiments and results

In this section, we evaluate the performance of the proposed method for mitosis detection on 2012 MITOSIS contest dataset and 2014 MITOSIS dataset. 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 released

publicly. 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. These images are HPFs at 40X magnification and are selected from breast cancer biopsy slides by pathologists. More than 300 mitotic cells are marked in this dataset and all pixels of each mitosis are annotated. We evaluate our method on the images produced by the Aperio XT scanner. The resolution of the scanner is $0.2456 \mu\text{m}$ per pixel. The HPF image has an area of $512 \times 512 \mu\text{m}^2$, 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 for testing. There are 226 and 101 mitotic cells 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 labeled mitotic cells. The test set contains 496 HPFs, and the number of mitotic cells is unknown since the ground truth annotations are withheld by the organizers. The mitotic cells are annotated by two pathologists. If there exists a conflict between them, another pathologist gives annotation and the final result is determined by the majority. As mentioned above, the annotation in this dataset only provides the single centroid pixel rather than all pixels for each mitotic cell.

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

Here we define some measures used for evaluating the accuracy of mitosis detection. *TP* is the number of detections that are ground truth mitosis, while *FP* is the number of detections that are not ground truth mitosis. The number of ground truth mitosis undetected is defined as *FN*. With these measures, we can calculate the *recall*, *precision*, and *F-score* using the following formulas:

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

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

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

4.2. Deep detection model on 2012 MITOSIS dataset

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

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, and finally train it for 20k iterations with the learning rate

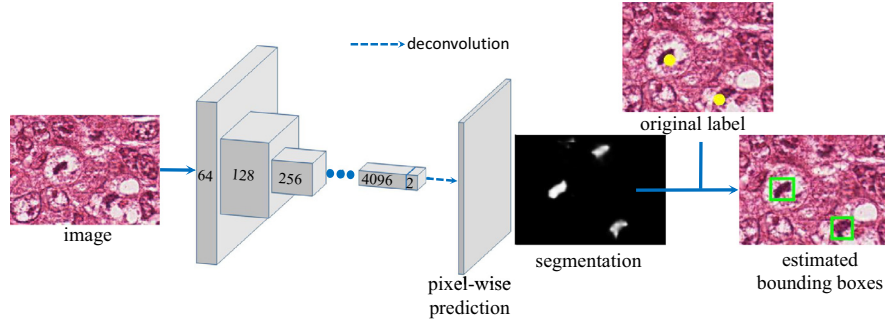


Fig. 6. The process of transforming centroid annotations to bounding box annotations by DeepSeg model. DeepSeg produces pixel-wise segmentation probability map. For each mitosis in ground truth, we locate a corresponding segmented mitotic region and generate a bounding box. The green boxes denote the estimated bounding boxes. The top segmented blob does not match a mitosis, so we ignore it. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

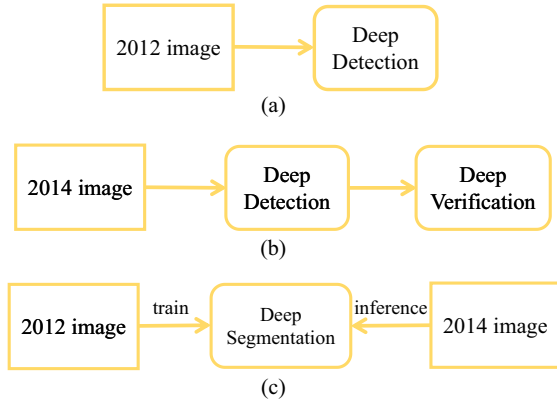


Fig. 7. The pipelines of the two mitosis dataset. (a) shows the detection pipeline of 2012 MITOSIS dataset, which merely utilizes the DeepDet model. (b) shows the detection pipeline of 2014 MITOSIS dataset, which needs the DeepVer model to refine the detections. (c) shows that the DeepSeg model is trained on 2012 MITOSIS dataset and then applied on 2014 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. 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

of 0.00001. We set momentum to 0.9, weight decay to 0.0005, and batch size to 1.

4.2.2. Data augmentation of training data

Even the TITAN X GPU has 12 GB memory, the full sized HPF image is still too large to be input into the DeepDet model. Thus we need to crop image patches from the full HPF images.

The detection model is initially designed for common objects, so mitosis under the original image scale is relatively small for the detector. Hence we enlarge the image patches before training. To seek the appropriate scale, we crop patches of different sizes from the histology images and uniformly re-scale them to 1024×1024 pixels. Then we use these patches with different scales 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 the original scale image patches. And the performance improves

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

drastically as the scale increases to 1.6. Further increasing the scale does not bring a distinct effect on the F-score. 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 these sampled image patches to 1024×1024 pixels.

It is noticeable that the mitotic cells in the boundary of patches may be split into two or more small parts. If the generated mitotic part is tiny, it should not be regarded as a valid mitotic region. Hence we remove the image patches that contain small cross-boundary mitotic cells from the training data. We set the area threshold to 800 pixels according to some evaluation experiments. The moderate threshold keeps appropriate positive samples as much as possible, and simultaneously filters out tiny mitosis parts which may interfere with the 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 full 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 DeepDet training, especially when the number of mitotic samples in the original dataset is very small.

Table 2 shows the performance of DeepDet models trained on different training sets. The number in brackets of the first column is the area threshold used to screen out the marginal mitosis. For instance, the “AreaTh(800)” training set removes the image patch containing a boundary mitotic region smaller than 800 pixels. The “+Rot” means that the training set includes rotated images; otherwise, its data augmentation only includes mirroring. We note that the image rotation improves the F-score remarkably, which demonstrates the importance of data augmentation. Fig. 8 shows the detection results of “AreaTh(800)+Rot” detector on some sample regions from 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.

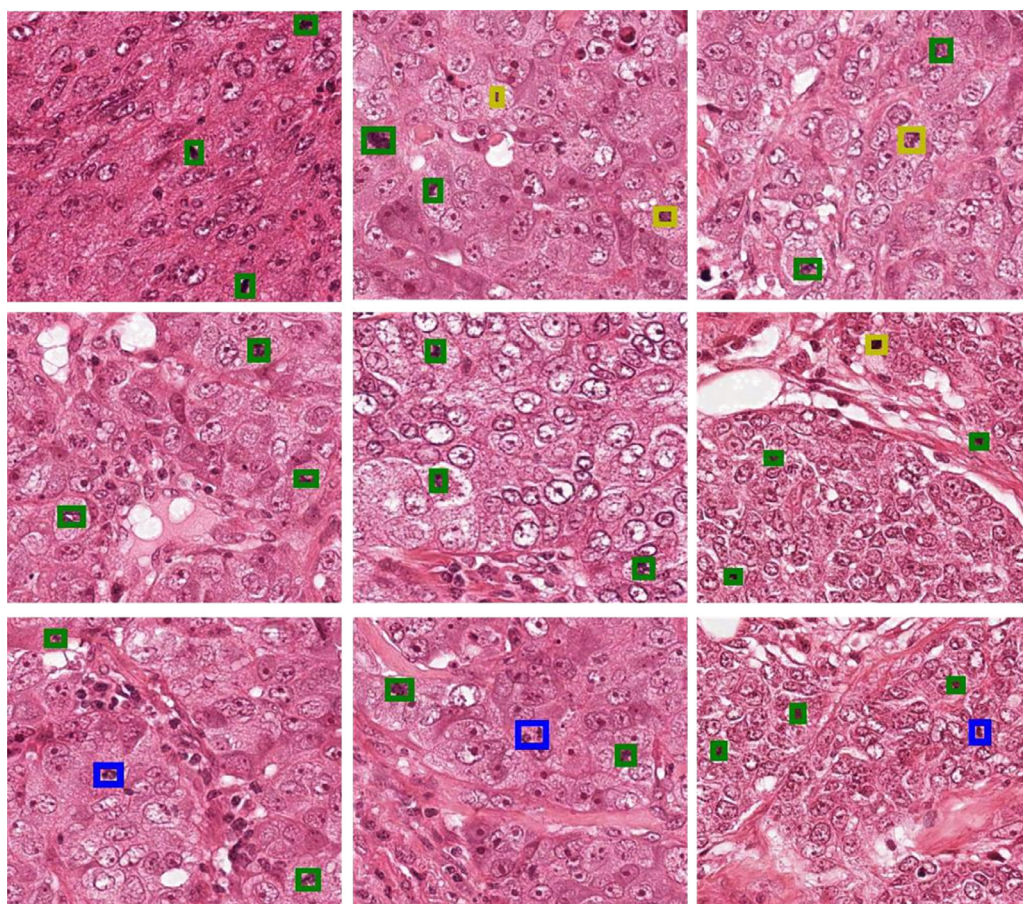


Fig. 8. Mitosis detection results of DeepDet on some sample regions of histopathological slides from 2012 MITOSIS test set. Yellow, blue and green boxes denote false positives, false negatives and true positives, respectively. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 3

The performance of DeepDet models trained 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

Anchor batch size: In our RPN training, the “Anchor batch size” denotes the mini-batch size of anchors from a single training image. For the mitosis data, the number of mitosis in a single training image is often low. The default batch size is 256, and we evaluate some smaller values, e.g. 128 and 32. As reported in Table 3, the results of using small batch size are comparable to the best results. It indicates that the anchor 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 default anchor scales in the model, which are 128, 256 and 512. Considering that the size of mitotic cells is usually small, the largest anchor scale 512 is not appropriate here, even though the images have been resized with scale 2. As shown in Table 3, the F-score improves by 2.2% when we replace the scale 512 with 64. The improvement indicates that

the anchor scale 64 is suitable for the detection. Note that using merely two scales of 64 and 128 also yields good results. It makes sense since most of mitotic cells range between 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 produced by RPN. These proposals are then judged by the following classification sub-network. The “proposal number” in Table 3 is the top-N. 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, the DeepDet 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 foreground object in this application, 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. We can regard the proposals produced by the RPN 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 simply remove its classification sub-network to get our RPN model. As the parameters of fully connected layers in the classi-

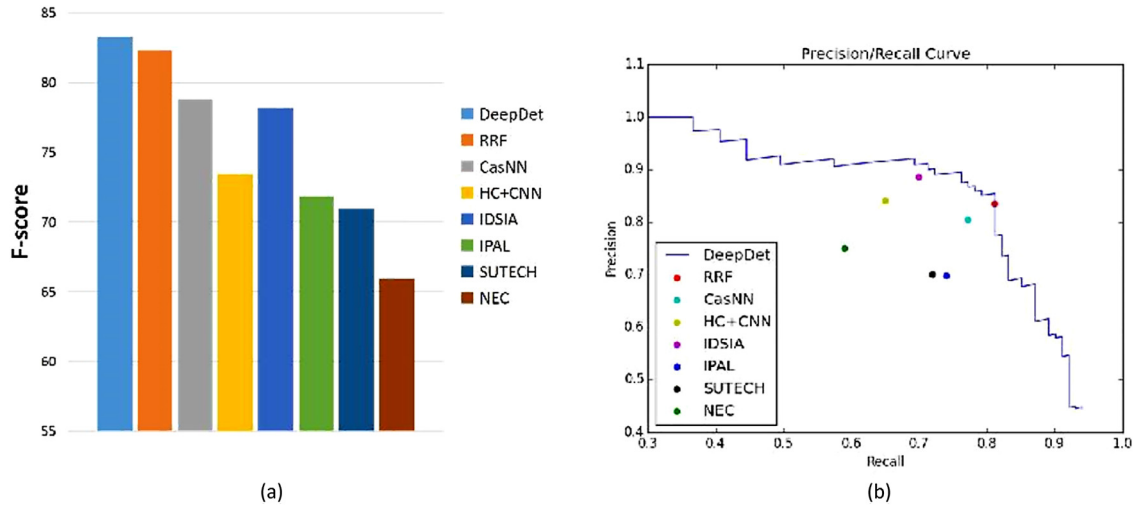


Fig. 9. (a) F scores of DeepDet and other methods on 2012 MITOSIS test set. (b) Performance comparison with others in the PR plane.

Table 4

Performance comparison of DeepDet with other competing approaches on 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

fication sub-network account for the majority of weights in the DeepDet, the size of the remaining RPN module is very small (about 30M bytes). The F-score of this RPN model on the 2012 MITOSIS test set is 0.796. The competitive performance suggests that the extracted RPN model can be a good detector for this single-category detection problem. However, when we train a new RPN with the same mitosis data independently, the F-score is 0.768. The nearly 3% gap indicates that the training mode of sharing convolutional features between the RPN and the classification sub-network can actually produce better convolutional features than training the RPN alone.

4.2.5. Comparison with other methods

We compare our method with some other approaches in this section. 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) and NEC (Malon et al., 2013) are the top four results participating in the Mitosis detection contest in ICPR 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 with handcrafted features, and other methods utilize convolutional features only.

4.3. Mitosis detection on 2014 MITOSIS dataset

In this weakly annotated dataset, we utilize the DeepSeg model to segment the histology image and estimate bounding boxes for mitoses. Then we train the DeepDet with these predicted annotations. Finally, to refine detections results and filter out false positives, we use the DeepVer model.

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

4.3.1. Implementation of DeepSeg model

As described above, the annotation of 2014 MITOSIS dataset only includes the centroid pixel for each mitotic cell. To achieve a finer bounding box annotation for training the detector, we utilize the DeepSeg model to segment the images and estimate a bounding box for each mitosis.

We use the precisely annotated 2012 MITOSIS dataset to train the DeepSeg model. We sample patches of 521×521 pixels from HPF images, and then mirror and rotate them to augment the training data. Since there are much more negative pixels than positive pixels, we perform more augmentation on patches containing positive pixels than patches only having negative pixels to balance the data.

The DeepSeg model is trained with Caffe framework (Jia et al., 2014). The FCN base model we used is trained on the 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 modify the channel number of prediction layer to two for predicting object or background.

There are three different versions in FCN: 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, and then train the finer stride versions of FCN models. FCN-16s model is initialized with the parameters of the FCN-32s, and is trained with a learning rate $1e-13$. 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 to produce segmentation map. Since the spatial resolution of a HPF is too large, we cut the HPF image to 16 patches evenly when testing, and then combine the segmentation outputs of these patches into the segmentation map of full image. We perform a binarization processing on the segmentation map. For each original centroid label, we locate a segmented mitosis blob that covers the centroid

¹ <https://github.com/BVLC/caffe/wiki/Model-Zoo>.

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(c)	0.559
DeepDet+Seg+Ver(f)	0.572

pixel. Then we get a bounding box to enclose the segmented blob as the new annotation. If there is no segmented mitotic region covering the centroid, we will assign a fixed bounding box to it. In our experiment, there are about 104 mitotic cells that do not have corresponding segmented blobs, accounting for 14% of all mitoses.

4.3.2. Implementation of DeepVer model

The estimated mitotic bounding box on 2014 MITOSIS dataset may be not very reliable, which can introduce inaccurate supervision during the DeepDet training. The detection results of DeepDet on this dataset are not very accurate, hence we exploit the DeepVer model to further verify them. 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 extract a 96×96 patch. For positive patch, we rotate it with a 90-degree step to produce more positive samples for balancing the data bias. Meanwhile, we crop patches from centroid ground truth, and perform the same rotation augmentation on them. Moreover, the original annotation has provided some mimics which are prone to be wrongly identified as mitosis. We add these samples to our training data to improve the model capacity of recognizing hard negatives. Finally, there are 16,248 training patches in total. We train a 50-layer ResNet as the DeepVer. We train the model for 12,000 iterations with a batch size of 64. The learning rate is 0.01 and momentum is 0.9, and weight decay is 0.0001.

4.3.3. The effectiveness of DeepSeg and DeepVer models

We conduct some ablation experiments to demonstrate the effectiveness of our deep segmentation model and deep verification model.

DeepDet (fixed): We first train a baseline 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. The anchor scales in RPN training are 64, 128 and 256. We assign a simple and uniform bounding box for each mitotic cell. The side length of the box is fixed to 40 pixels. The originally marked centroid pixel lies in the center of the box. Through a careful validation on the dataset, the 40×40 bounding box is found to be the best fixed annotation. We name the detector trained with fixed bounding box annotation as “DeepDet (fixed)”.

DeepDet+Seg: This DeepDet model is trained with the same data and parameter setting as the DeepDet (fixed), but it utilizes the estimated bounding box annotations produced 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 indicates that the refined annotations from DeepSeg model are conducive to the DeepDet training.

DeepDet+Seg+Ver: We then apply the DeepVer model to the detection results of DeepDet+Seg. Here we take two strategies to exploit the DeepVer model. The first method is a cascade way that simply takes the DeepVer classification results on the detection patches as the final decisions. We name this method as “DeepDet+Seg+Ver(c)”, as it utilizes the three deep models and the

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

“c” stands for “cascade”. In this experiment, the DeepVer only processes the detected patches with DeepDet+Seg score higher than 0.94. The DeepVer model can filter out many false positives of DeepDet+Seg and hence improves the F-score to 0.559 as shown in Table 5. The second method is to combine the confidences of the detection model and verification model. As described in Eq. (1), the fusion takes a weighted sum of the scores of the two models. The weight ω is optimized on the validation set, and is 0.9 in our experiment. We name this method as “DeepDet+Seg+Ver(f)”, in which the “f” stands for “fusion”. It achieves a 0.572 F-score on the validation set. Compared with the DeepDet+Seg+Ver(c) method, the fusion method can take advantage of the two prediction scores adequately and achieves a superior detection quality. Thus we only use the fusion strategy in the following experiments.

Fig. 10 illustrates some detections comparisons between the DeepDet+Seg and DeepDet+Seg+Ver on the 2014 MITOSIS validation set. The first nine columns confirm that the DeepDet+Seg+Ver can effectively identify false positives of DeepDet, due to the strong discriminative capacity of DeepVer model. The last two columns illustrate mitotic cells missed by DeepDet+Seg are identified by the DeepDet+Seg+Ver. Fig. 11 shows the detection results of proposed DeepDet+Seg+Ver approach on four HPFs from the 2014 MITOSIS validation set.

4.3.4. Results on the test set

We now 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 has been released by the organizers. Here we train the verification model on the detection patches of all training data (including the validation set).

Experimental results on test set are shown in Table 6. The DeepDet+Seg achieves a 0.398 F-score, which is a comparable result to the state-of-the-art as we will see below. We choose a high score 0.984 as the decision threshold in DeepDet+Seg, so its recall is relatively low. This high threshold is a trade-off between the precision and recall. For example, if we reduce the detection threshold score to 0.9, the recall will increase to 0.522, but the precision will drop to 0.192, leading to a much worse F-score 0.281. Compared with the DeepDet+Seg, DeepDet+Seg+Ver has an identical precision, but a significant improvement in recall, which increases the F-score to 0.437. We take the DeepDet+Seg detections with scores higher than 0.88 for the following verification phase. The relatively low threshold of detection model involves more true detections, and simultaneously the verification model can remove the false positives effectively, so the F-score is remarkably higher than the DeepDet+Seg.

4.3.5. Comparison with other methods

The performance comparison of the proposed method with other approaches is reported in Table 7. Our method achieves a state-of-the-art performance with F-score 0.437, outperforming all other methods except for the CasNN (Chen et al., 2016a). 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% lower F-score.

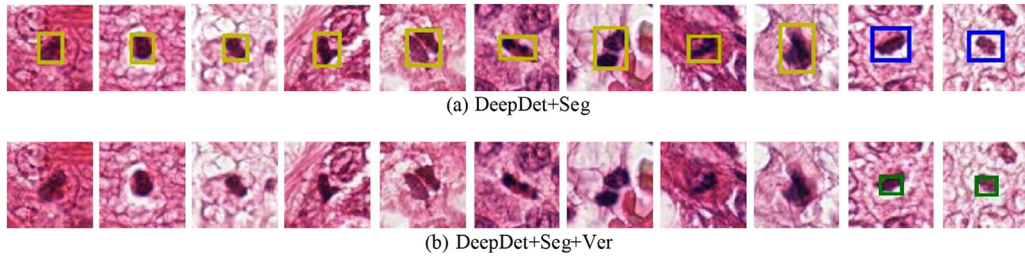


Fig. 10. Example detections of the proposed methods on the validation set of 2014 MITOSIS dataset. (a) shows the detection results of DeepDet+Seg. (b) shows the results of DeepDet+Seg+Ver(f), which combine the predictions of detection model and verification model. We show eleven examples to compare the differences between the two results. Yellow, blue and green boxes denote false positives, false negatives and true positives, respectively. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

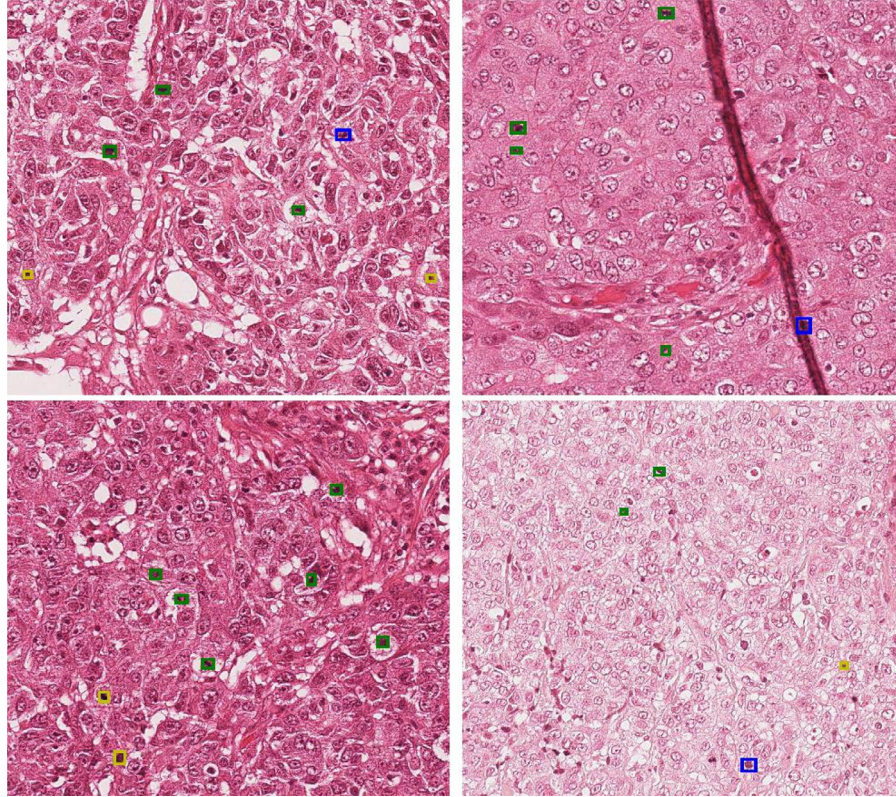


Fig. 11. Mitosis detection results of DeepDet+Seg+Ver on four HPFs of the 2014 MITOSIS validation set. Yellow, blue and green boxes denote false positives, false negatives and true positives, respectively. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

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

4.4. Discussion

It is interesting that our method achieves a better performance than CasNN (Chen et al., 2016a) on 2012 MITOSIS dataset, but an inferior performance on 2014 MITOSIS dataset. The reason may be that our method is based on a detection architecture which

requires accurate bounding box ground truth, while the CasNN is based on a classification model that does not demand such annotation. The centroid format annotation of 2014 MITOSIS dataset is appropriate for a classification model, but not for a region based detection model. Even so, with the aid of estimated bounding box label, we still achieve a high performance on 2014 dataset. Furthermore, the performance of our method could be enhanced impressively if we have better supervision, like the annotation in 2012 dataset.

For the consistency of our proposed system, we also apply the DeepVer model on the 2012 MITOSIS dataset. The fusion on scores of DeepDet and DeepVer models yields a 0.817 F-score. Unfortunately, using a DeepVer model does not improve the performance on this dataset. The main reason is that the discriminative capacity of the DeepDet model trained on 2012 dataset is sufficiently powerful. The precision of the detection model is already very high so that it is hard to improve further. The target of the verification model is to remove the false positives produced by detection model. Now that the false positives are few, the effect

of DeepVer is minor. Moreover, the number of detections in this dataset is very low, making it hard to train a deep classification model effectively. In summary, we think it is unnecessary to apply the DeepVer on 2012 MITOSIS dataset.

4.5. Time analysis

The target of automatic mitosis detection is to help pathologists in clinical applications. Since the number of HPFs in a single whole slide may be huge, the speed of detection algorithm is very critical. For a HPF with 2084×2084 pixels in 2012 MITOSIS dataset, the DeepDet takes about 0.7 s to perform detection and the RPN takes about 0.66 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 detection model takes 0.35 s per 1539×1376 HPF. The elapsed time of verification model depends on the processing speed on an image patch and the number of candidate patches DeepDet+Seg produced. The DeepVer model takes 0.023 s for each patch, and the DeepDet+Seg model averagely produces about two detections with score higher than 0.8 in a HPF. Hence the verification model takes about 0.05 s per HPF. In total, the DeepDet+Seg+Ver system takes about 0.4 s for a HPF from 2014 MITOSIS dataset.

IDSIA (Cireşan et al., 2013) requires 31 s to apply a network on an input HPF image. For a better performance, it utilizes two networks on eight image variants, which takes 8 min. Our approach is significantly faster than IDSIA. CasCNN (Chen et al., 2016a) takes about 0.5 s to process per 4M pixels HPF (size 2084×2084) from 2012 MITOSIS dataset. Our method is comparable to it in speed.

5. Conclusions and future works

In this paper, we propose a system named DeepMitosis for mitosis detection in H&E stained slide images. We adopt a general object detection method to the histopathology images and achieve 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 our 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 the 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.

Our mitosis detection method is performed in HPFs pre-selected by pathologists. In addition, the detection results need to be verified by the pathologists for a better accuracy, and the tumor proliferation score need to be predicted according to the mitotic count. So the whole cancer grading system is still a semi-automatic method, requiring the effort from human. In practical scenario, a unified and automatic cancer grading system makes more sense. It includes not only mitosis detection, but also the processes of selecting regions of interest and predicting the tumor proliferation score from the detection results. Hence it is meaningful to build a complete cancer grading system based on our mitosis detection approach.

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. 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. In addition, we will explore how to build a complete system to predict tumor proliferation score from the whole slide images directly.

Acknowledgments

This work is supported by the National Natural Science Foundation of China (Grant nos. 61503145 and 61572207), the Young Elite Sponsorship Program by CAST (No. YESS 20150077), and the Program for HUST Academic Frontier Youth Team. This work is further supported by the National Science Foundation Grant IIS-1302164 and the CCF-Tencent RAGR (No. 20170104).

References

- Chatfield, K., Simonyan, K., Vedaldi, A., Zisserman, A., 2014. Return of the devil in the details: delving deep into convolutional nets. 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 13th AAAI Conference on Artificial Intelligence. AAAI Press, pp. 1160–1166.
- Chen, H., Wang, X., Heng, P.-A., 2016b. Automated mitosis detection with deep regression networks. In: Proceedings of 2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI). IEEE, pp. 1204–1207.
- Christ, P. F., Ettlinger, F., Grün, F., Elshaera, M. E. A., Lipkova, J., Schlecht, S., Ahmaddy, F., Tatavarty, S., Bickel, M., Bilic, P., et al., 2017. Automatic liver and tumor segmentation of CT and MRI volumes using cascaded fully convolutional neural networks. arXiv:1702.05970.
- 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. Springer, pp. 411–418.
- 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 (5), 403–410.
- 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.
- Havaei, M., Davy, A., Warde-Farley, D., Biard, A., Courville, A., Bengio, Y., Pal, C., Jodoin, P.-M., Larochelle, H., 2017. Brain tumor segmentation with deep neural networks. *Med. Image Anal.* 35, 18–31.
- 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: Proceedings of 2012 21st International Conference on Pattern Recognition (ICPR). IEEE, pp. 1856–1859.
- Irshad, H., et al., 2013. Automated mitosis detection in histopathology using morphological and multi-channel statistics features. *J. Pathol. Inform.* 4 (1), 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:1408.5093.
- Kamnitsas, K., Ledig, C., Newcombe, V.F., Simpson, J.P., Kane, A.D., Menon, D.K., Rueckert, D., Glocker, B., 2017. Efficient multi-scale 3d CNN with fully connected crf for accurate brain lesion segmentation. *Med. Image Anal.* 36, 61–78.
- 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: Proceedings of 2012 21st International Conference on Pattern Recognition (ICPR). IEEE, pp. 149–152.
- Krizhevsky, A., Sutskever, I., Hinton, G.E., 2012. ImageNet classification with deep convolutional neural networks. In: Proceedings of 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: Proceedings of International Conference on Artificial Intelligence and Statistics (AISTATS), 2, p. 5.
- Li, C., Wang, X., Liu, W., 2017. Neural features for pedestrian detection. *Neurocomputing* 238, 420–432.
- Li, J., Liang, X., Shen, S., Xu, T., Feng, J., Yan, S., 2015. Scale-aware fast R-CNN for pedestrian detection. 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: Proceedings of European Conference on Computer Vision. Springer, pp. 740–755.
- 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. *J. Pathol. Inform.* 4 (1), 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. *IEEE Trans. Image Process.* 14 (9), 1360–1371.
- 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*. Springer, pp. 94–102.
- Paul, A., Mukherjee, D.P., 2015. Mitosis detection for invasive breast cancer grading in histopathological images. *IEEE Trans. Image Process.* 24 (11), 4041–4054.
- Ren, S., He, K., Girshick, R., Sun, J., 2015. Faster R-CNN: Towards real-time object detection with region proposal networks. In: *Proceedings of Advances in Neural Information Processing Systems*, pp. 91–99.
- Ronneberger, O., Fischer, P., Brox, T., 2015. U-net: convolutional networks for biomedical image segmentation. In: *Proceedings of International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, pp. 234–241.
- Roth, H.R., Lu, L., Farag, A., Sohn, A., Summers, R.M., 2016. Spatial aggregation of holistically-nested networks for automated pancreas segmentation. In: *Proceedings of International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, pp. 451–459.
- 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. *J. Pathol. Inform.* 4, 4–8.
- 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. *Int. J. Comput. Vis.* 115 (3), 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. *Int. J. Comput. Vis.* 115 (3), 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: *Proceedings of European Conference on Computer Vision – ECCV 2012*. Springer, pp. 746–760.
- Sommer, C., Fiaschi, L., Hamprecht, F.A., Gerlich, D.W., 2012. Learning-based mitotic cell detection in histopathological images. In: *Proceedings of 2012 21st International Conference on Pattern Recognition (ICPR)*. IEEE, pp. 2306–2309.
- Tang, P., Wang, X., Huang, Z., Bai, X., Liu, W., 2017. Deep patch learning for weakly supervised object classification and discovery. *Pattern Recognition*, 71, 446–459.
- 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: *Proceedings of 2013 5th Conference on Information and Knowledge Technology (IKT)*. IEEE, pp. 406–410.
- Tek, F.B., et al., 2013. Mitosis detection using generic features and an ensemble of cascade adaboosts. *J. Pathol. Inform.* 4 (1), 12.
- TUPAC16, 2016. Tumor-proliferation-assessment-challenge. <https://http://tupac.tue-image.nl/> (Online; accessed 17.07.30).
- Uijlings, J.R., Van De Sande, K.E., Gevers, T., Smeulders, A.W., 2013. Selective search for object recognition. *Int. J. Comput. Vis.* 104 (2), 154–171.
- Veta, M., van Diest, P., Pluim, J., 2013. Detecting mitotic figures in breast cancer histopathology images. In: *SPIE Medical Imaging. International Society for Optics and Photonics*, p. 867607.
- 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. *Med. Image Anal.* 20 (1), 237–248.
- Wang, H., Cruz-Roa, A., Basavanahally, 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. International Society for Optics and Photonics*, p. 90410B.
- Wang, X., Yan, Y., Tang, P., Bai, X., Liu, W., 2018. Revisiting multiple instance neural networks. *Pattern Recognit.* 74, 15–24.
- 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: *Proceedings of European Conference on Computer Vision*. Springer, pp. 443–457.