CHS-NET: A Cascaded Neural Network with Semi-Focal Loss for Mitosis Detection

Yanbo Ma[1]   MAYANBO@PKU.EDU.CN
Jiarui Sun[2]   1701120005@PKU.EDU.CN
Qiuhao Zhou[1]   ZHOUQH1994@PKU.EDU.CN
Kaili Cheng[1]   CHENGKAILI@PKU.EDU.CN
Xuesong Chen[1]   CEDARCHEN@PKU.EDU.CN
Yong Zhao [*] [1]   YONGZHAO@PKUSZ.EDU.CN

[1] *School of Electronic and Computer Engineering, Shenzhen Graduate School of Peking University, Shenzhen, China*

[2] *School of Electronic Engineering and Computer Science, Peking University, Beijing, China*

Editors: Jun Zhu and Ichiro Takeuchi

Abstract

Counting of mitotic figures in hematoxylin and eosin(H&E) stained histological slide is the main indicator of tumor proliferation speed which is an important biomarker indicative of breast cancer patients' prognosis. It is difficult to detect mitotic cells due to the diversity of the cells and the problem of class imbalance. We propose a new network called CHS-NET which is a cascaded neural network with hard example mining and semi-focal loss to detect mitotic cells in breast cancer. First, we propose a screening network to identify the candidates of mitotic cells preliminary and a refined network to identify mitotic cells from these candidates more accurately. We propose a new feature fusion module in each network to explore complex nonlinear predictors and improve accuracy. Then, we propose a novel loss named semi-focal loss and we use off-line hard example mining to solve the problem of class imbalance and error labeling. Finally, we propose a new training skill of cutting patches in the whole slide image, considering the size and distribution of mitotic cells. Our method achieves 0.68 F1 score which outperforms the best result in Tumor Proliferation Assessment Challenge 2016 held by MICCAI.

Keywords: Mitosis detection, Cascaded neural network, Hard example mining, Semi-focal loss

1. Introduction

Breast Cancer is the most common diagnosed cancer among women, and the second leading cause of death worldwide. Tumor proliferation rate is an important biomarker for breast cancer (Diest et al., 2004). To evaluate the growth rate of tumor, the pathologists count the number of mitotic cells in an area of 2 mm^2, which corresponds to 8-10 microscope high power fields (HPFs) (Veta et al., 2015). The mitosis count is part of the modified

---
[*]

Bloom-Richardson grading system (Elston and Ellis, 1991) and it is also considered as an independent prognostic factor (Baak et al., 2005).

Mitotic cell detection is a very challenging task for the following reasons. First, the WSI is examined under a microscope by a pathologist traditionally, hence mitosis count is a subjective procedure which is prone to discrepancies between observers. Second, WSI generally have thousands of single cells and mitotic cells also have different shape characteristics. In particular, two unformed daughter cells are produced at the telophase stage of mitosis, but we still consider them as one mitotic cell. This increases the difficulty of mitosis detection task. Third, mitotic cells can be very similar to other objects in the tissue, such as apoptotic cells and lymphocytes, leading to a lot of false positives in the detection process.

Thanks to the recent advances in deep learning, automated methods have been successfully adopted to the medical diagnostic imaging field. Generally, researchers use a deep neural network(DNN) to detect mitotic cells (Cireşan et al., 2013). However, the imbalance between positive and negative samples is a common problem in the medical imaging field and only one DNN is hard to solve it. Chen et al. (2016) figures out cascaded networks to alleviate class imbalance. From this, we add a novel feature fusion module in the similar structure like Resnet (He et al., 2015) in cascaded networks to detect mitotic cells, which can further improve the result. Other methods, such as online hard example mining(OHEM) (Shrivastava et al., 2016) and focal loss (Lin et al., 2017) are performed to maintain a manageable balance between the positive and negative samples. However, OHEM completely discards easy examples and focal loss cannot deal with the problem of error labeling in the datasat. In order to solve similar problems in face recognition, Facenet (Schroff et al., 2015) selected semi-hard negative samples since the hardest negatives can lead to bad local minima early on in training. Inspired by this, we propose semi-focal loss to ensure that excessive losses can be reduced to a certain threshold. We propose the offline hard example mining and add simple samples into the refined network in addition to the false positives to make up for the deficiency of OHEM. It is difficult to train the entire image directly as an HPF has 2000×2000 pixels and researchers generally cut patches of positive and negative samples by calculating IoU((Intersection-over-Union). Considering the location and size of mitotic cells, We propose a new method for obtaining patches which can consider both background and contextual information in the mitosis detection task.

TUPAC16 is a tumor proliferation assessment challenge and the third task is mitosis detection. We use the dataset provided by this competition to deal with mitosis detection problem.

The major contributions of this paper are as follows:

- We propose a new mitosis detection network called CHS-NET and a novel feature fusion module to improve the accuracy of detection.

- We design an effective method to alleviate the problem of class imbalance and error labeling. We use offline hard example mining to re-sampling the hard negative samples and improve the proportion of the false negatives in the dataset. Semi-focal loss is proposed to solve the problem of error labeling.

- We come up with a novel method for obtaining patches by calculating IoA(Intersection-over-Area).

2. Related Work

2.1. Machine Learning Methods

Machine learning has been successfully applied to mitosis detection. Vink et al. (2013) proposes two nucleus detectors with the basis of AdaBoost to extract different features and uses colour deconvolution. This method can detect the boundary of the nucleus more effectively and the detection rate was 95%. Irshad (2013) proposes an automated mitosis detection framework based on different multi-channel statistical features with morphological features by using decision tree classifier.

2.2. Deep Learning Methods

In order to solve the problem that the number of negative samples is too large, Cireşan et al. (2013) uses the probability that a negative sample is wrongly considered as a positive sample as the sampling probability of the negative samples, thereby increasing the proportion of negative samples that are easily considered as positive samples in training data. Chen et al. (2016) proposes the use of two neural networks for cell classification. The first network can filter out a large number of negative samples that can be easily distinguished in the data. This network has fewer parameters to prevent overfitting. The second neural network categorizes the preliminary screening data and further screens out negative samples that are difficult to distinguish. This paper also uses Imagenet's pretrained model to have better result through migration learning. The first place in the TUPAC 2016 competition (Paeng et al., 2017) uses the ResNet32 network to increase accuracy by increasing the number of false positives. To solve the problem of class imbalance, this article uses two step training. The first network picks out 100,000 examples that are mistakenly considered as positive samples. They add these false positives to data sets and then obtain a second neural network. Zerhouni et al. (2017) uses the online hard-to-negative sample mining method to update false positive samples in a timely manner so as to increase the proportion of false positive samples in training data in a timely manner. First, a neural network is used to select false positive samples to form a new sample set. Then, equal amounts of data are randomly selected in this set, positive sample set and true negative set as the current round of training data. Each iteration is updated according to this method until the network converges.

In the biomedical field, it is difficult to obtain a large amount of training data. Bayramoglu and Heikkilä (2016) compare four different CNN models and demonstrates that transfer learning can alleviate this problem. The results show that the pre-trained initialization network parameters can improve the classification performance of any model, and it is easy to converge. Many new network structures have also been proposed. Sirinukunwattana et al. (2016) proposes a Spatially Constrained Convolutional Neural Network (SC-CNN) to perform nucleus detection. SC-CNN calculates the probability of a pixel being the center of the nucleus, where the high probability value is constrained in space to locate near the core center. Albarqouni et al. (2016) introduces a novel concept for learning from crowds. He proposes a new multi-scale CNN AggNet which is designed to directly handle data aggregation through additional crowdsourcing and generates a basic fact tag from the notes of non-experts.

Ma[1] Sun[2] Zhou[1] Cheng[1] Chen[1] Zhao

3. Our Method

3.1. Network

In this section, we will describe the mitotic cell detection neural network named CHS-NET that we propose based on the characteristics of mitotic cell detection tasks in detail.

By analyzing the high power fields(HPFs), we easily find that the proportion of mitotic cells in the entire WSI is extremely small and the characteristics are relatively single. Non-mitotic cells, including apoptotic cells, take up a very large portion of WSI and vary in shape. So class imbalance is a serious problem. The non-mitotic cells and the non-cellular fractions in the WSI have obvious differences in appearance characteristics. Some of these cells can be identified as non-mitotic cells clearly, while others are very similar to mitotic cells, as shown in Fig. 2. Therefore, how to train and obtain a network that can distinguish mitotic cells and non-mitotic cells effectively, especially identify non-mitotic cells that are very similar to mitosis is the focus of this article.

Based on the above analysis, this paper proposes CHS-NET which is a multi-task cascaded neural network to detect mitotic cells. Our multi-tasking cascaded mitotic cell detection network mainly consists of two parts as shown in Fig. reffig:frame.eps
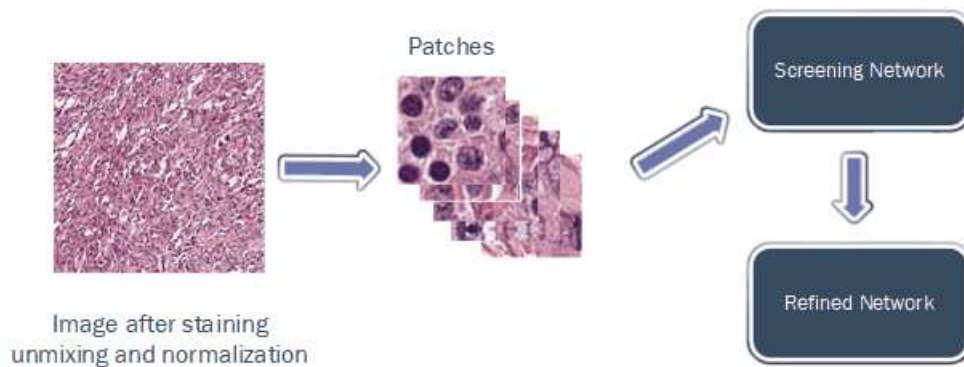

Figure 1: An overview of the CHS-NET.

3.1.1. Screening Network

The goal of the screening network is to screen out the simple negative samples and ensure that the recall rate of mitotic cells is nearly 100%. We can see that the negative samples in the first row of Fig. 2 are very different from the positive samples and are very easy to distinguish by a simple neural network, but some hard examples are very troublesome to the neural network. Therefore, the challenge in the detection of mitotic cells is how to make the neural network focus on distinguishing between difficult samples and mitotic cells. A large number of simple examples are easy to acquire the attention of the neural network inevitably. We also need the neural network to be able to identify simple examples. Therefore, our screening network can identify mitotic cells, simple negative samples, and negative samples that are easily misclassified roughly. It consists of a feature extractor (Table. 1) and a perceptron with multi feature fusion layers(Fig. 3).
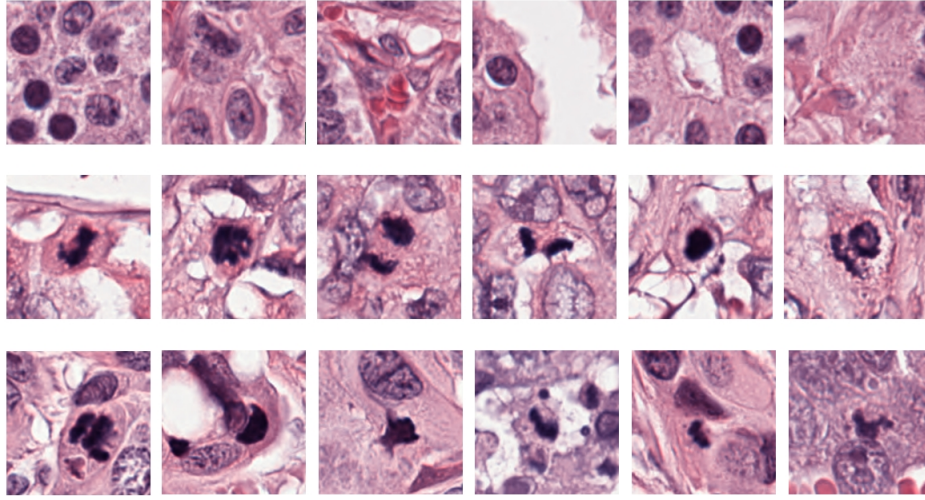
Figure 2: The patches of mitotic cells, non-mitotic cells and some hard samples. The patches of first row are easily distinguishable non-mitotic cells and the patches of second row are mitotic cells. The patches of third row are some non-mitotic cells that are not easily distinguished. The size of patch is 128×128.
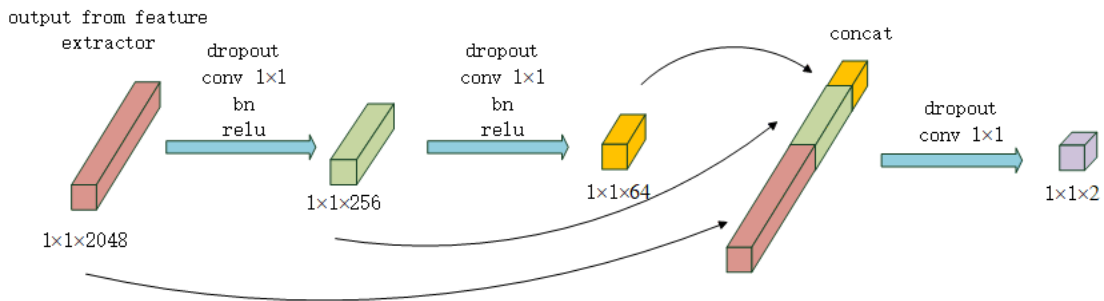

Figure 3: The structure of feature fusion in perceptron.

The screening network classifies each of the patches that are input into the neural network to distinguish between mitotic cells and non-mitotic cells. We set the label of mitoses to 0 and the others to 1. Each patch in the screening network will get the probability of mitosis. In general, the patch will be considered as a mitosis if the probability is greater than 0.5. However, we need to ensure that the positive samples' recall rate is nearly 100% and we have to screen out simple negative samples simultaneously. In our experiment, we set the threshold to 0.2 and select negative samples with the probability greater than the threshold as false positives. After this method, the screening network will identify a large number of simple samples, and almost no mitotic positive samples will be missed. The specific process and experimental results will be detailed in Section 4.

Table 1: The structure of feature extractor. The conv2_x to conv6_x are the basic residual blocks.

Name	Layer setting	Output Size
input		H×W
conv1	7×7, 128, stride 1	H×W
conv2_x	$\begin{pmatrix} 1\times1,\ 32 \\ 3\times3,\ 32 \\ 1\times1,\ 128 \end{pmatrix}$ ×3, stride 1	H×W
conv3_x	$\begin{pmatrix} 1\times1,\ 64 \\ 3\times3,\ 64 \\ 1\times1,\ 256 \end{pmatrix}$ ×3, stride 2	$\frac{1}{2}$H×$\frac{1}{2}$W
conv4_x	$\begin{pmatrix} 1\times1,\ 128 \\ 3\times3,\ 128 \\ 1\times1,\ 512 \end{pmatrix}$ ×5, stride 2	$\frac{1}{4}$H×$\frac{1}{4}$W
conv5_x	$\begin{pmatrix} 1\times1,\ 256 \\ 3\times3,\ 256 \\ 1\times1,\ 1024 \end{pmatrix}$ ×3, stride 2	$\frac{1}{8}$H×$\frac{1}{8}$W
conv6_x	$\begin{pmatrix} 1\times1,\ 512 \\ 3\times3,\ 512 \\ 1\times1,\ 2048 \end{pmatrix}$ ×3, stride 2	$\frac{1}{16}$H×$\frac{1}{16}$W
	bn+relu+average pool	1×1

3.1.2. Refined Network

The main role of refined network is to complete the second task, which will classify the patch accurately after the preliminary screening network. Its structure is the same as screening network's. However, it needs more train to distinguish mitotic cells and other things.

For ordinary cascaded network, we use the sample with the probability greater than the threshold in the screening network as the input for the refined network. A large number of negative samples can be filtered by the screening network. The refined network can make the neural network focus more on negative samples which are not easily identifiable. Each patch in the refined network will also get the probability of mitosis and we set the threshold to 0.85. Each patch will get two probability by the screening network and the refined network in the test. Only when the two probability of the patch can reach the threshold of screening network and the threshold of refined network respectively, this patch will be considered as a positive sample. As a large number of negative samples are filtered by the screening network, the cascaded network can alleviate the problem of class imbalance. We introduce the hard example mining in Section 3.2 to improve the result.

3.1.3. Feature Fusion

In this paper, we will convert the mitotic cell detection task into binary classification task for patches. Deep learning classification networks are generally divided into two parts: one

is feature extractor and the other is the classification layer. A similar approach is used in our CHS-NET.

In the feature extraction layer, we use a structure similar to ResNet (He et al., 2015). The structure of ResNet can prevent network overfitting and the phenomena of gradient dispersion. Inspired by the work of GoogLeNet (Szegedy et al., 2015), ResNet and others, we have a global pooling average operation on the feature maps out of the feature extraction layer to further reduce over-fitting and improve accuracy. We introduce two fully connection layers in the feature fusion part. The fully connected layer acts as a classifier throughout the convolutional neural network.

Inspired by PixelNet (Bansal et al., 2017) and DenseNet (Huang et al., 2016), we do not use single-line full-connectivity to extract network features. As shown in Fig. 3, we fuse the feature vectors of each layer and then we do the final classification. Multi-level feature fusion has several advantages. First, the multi-level feature fusion can make the feature vector of each layer connect to the loss function layer directly, avoiding the situation that the gradient dispersion causes the network not easy to converge. It is necessary to introduce multi-level vector fusion, because there are a large number of parameters in the fully connected layer and gradient diffusion is likely to occur. Second, feature vectors in different dimensions contain different information. The introduction of multi-level feature fusion is conducive to explore more nonlinear predictors that improve overall accuracy.

3.2. Hard Example Mining

Hard example mining is a method of re-sampling the hard negative samples so that the neural network pays attention to the false positive feature, thereby improving the classification effect. From the foregoing, it can be found that there are great class imbalance problems in mitosis detection task. Then the precision of the positive samples is:

$$precision = \frac{tp}{tp + fp} \tag{1}$$

For positive samples, the number of true positives is tp, the number of false positives is fp, and the number of false negatives is fn. The recall rate of the positive samples is:

$$recall = \frac{tp}{tp + fn} \tag{2}$$

Our final assessment method is to calculate the F1 score:

$$F1\ score = \frac{2 \times precision \times recall}{precision + recall} \tag{3}$$

In order to get a higher F1 score, we need to reduce the number of false positives. Therefore, it is necessary to reduce the number of false positives by hard example mining. Hard example mining, a bootstrapping method, trains a model with an initial subset of negative examples, and then collects negative examples that are classified by this initial model incorrectly to form a set of hard negatives. We use this method to filter out false positives as a new set in the cascaded network. Through experiment, we have found that in the refined network, if the proportion of positive samples, negative samples, and false positives is 1:1:3, the network will make the neural network more focused on false positives and more accurately distinguish mitotic cells.

3.3. Semi-Focal Loss

There are widespread class imbalance problems in medical data sets. Firstly, since most of them are simple and easily distinguishable negative samples, the training process cannot fully learn information of positive samples. Secondly, these simple negative samples still produce a certain amount of loss, and the large quantity will play a major role in loss. Therefore, they will dominate the gradient update direction and obscure important information.

To solve the problem of class imbalance, Kaiming He's team propose focal loss (Lin et al., 2017) to improve the loss function. We define p_t for binary classification, p is the model estimated probability for mitotic cells with label $y = 0$:

$$p_t = \begin{cases} p & \text{if } y = 0 \\ 1 - p & \text{otherwise} \end{cases} \tag{4}$$

The focal loss(FL) is:

$$FL(p_t) = -\alpha(1 - p_t)^\gamma log(p_t) \tag{5}$$

α is a weighting factor to address class imbalance. $-log(p_t)$ is the cross entropy (CE) loss for binary classification. $(1 - p_t)^\gamma$ is a modulating factor to the cross entropy loss. The purpose is to make the model more focus on difficult-to-classify samples during training by reducing the weight of easy-to-classify samples. We set tunable focusing parameter $\gamma \geq 0$. In our model, α is 5 and γ is 2.

We analyze the result of a cascaded network with hard example mining, and find that some negative samples that are mislabeled as positive samples are indeed very similar to positive samples. Moreover, identifying all mitotic cells in a WSI is very difficult for a pathologist. Therefore, there may be negative samples with incorrect label in the dataset.

We propose the semi-focal loss to solve the problem of error labeling. The loss of mislabeled mitotic cells can be very large, and this will affect the convergence of the network. So, if $(1 - p_t)^\gamma > threshlod$, we replace $(1 - p_t)^\gamma$ with $threshold$. The formula is:

$$FL(p_t) = -\alpha log(p_t)min((1 - p_t)^\gamma, threshold) \tag{6}$$

By using smi-focal loss, we alleviate the problem of sample labeling errors.

3.4. Data Process

3.4.1. PATCH SELECTED AND LABELED

The input of screening and refined network is patch-level images. The corresponding output is the label of each patch. In the test, we hope one to four patches can be mapped to each mitotic cell through a refined network. The final result can be obtained by non-maximum suppression(NMS). Therefore, we need to divide the original images into several mitotic patches and other patches based on the location of the mitotic cells in training. By observing the mitotic cell image, we can find that the size of mitotic cells is fixed basically. So, mitosis detection task can use the same size box. A patch with the size of 64×64 can cover almost a entire mitotic cell. But the information surrounding the mitotic cells is helpful to the neural network to determine if this is a mitotic cell. So the patch size we finally selected

is 128×128. The general target detection task calculates the IoU between positive sample and the newly selected patch in the selection of negative samples. If the IoU is less than a threshold, it is treated as a negative sample. Because mitotic cells are located in the center of the patch generally, we use a new method to select negative samples. Our input is the patch resized to 64×64 and its center part with same size.

As shown in Fig. 4, the blue box is a positive sample with the size of 128×128. The green box is smaller than the blue box with the size of 64×64. It can contain the whole mitotic cell. The yellow box is a newly selected patch and its size is 128×128. Now we want to determine if this box is positive or negative. The general method is to calculate the IoU of the yellow box and the blue box.

$$IoU = \frac{area(blue\ box) \cap area(yellow\ box)}{area(blue\ box) \cup area(yellow\ box)} \tag{7}$$

Considering the position of mitotic cells in the patch, we have defined a new parameter named IoA. If the IoA is less than a threshold, the yellow box is a negative sample. In this paper, the threshold is 0.0625.

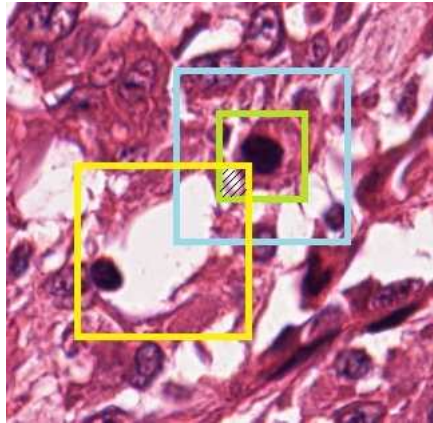$$IoA = \frac{area(green\ box) \cap area(yellow\ box)}{area(yellow\ box)} \tag{8}$$

Figure 4: The IOA calculation. The blue box is a positive sample with the size of 128×128. The green box is smaller than the blue box with the size of 64×64. It can contain the whole mitotic cell. The yellow box is a newly selected patch and its size is 128×128. we consider the yellow box and the green box when calculating IOA.

Specifically, we take the coordinate of the mitotic cell as the center point and obtain a patch with the size of 128×128 from the original images. We can get a positive sample set marked as S1 after processing all mitotic cells. For each patch in S1, we keep the same center point and extract patches with the size of 64×64. These patches constitute set S2. A box with the size of 128×128 traverses the full image from the upper left corner with a stride of 64 in each images. We will get a lot of patches and they form a set S3. A mitotic

cell may be contained by multiple patches. So , we use the NMS to remove duplicated boxes at the time of the test. For a patch in S3, if the IoA between it and each patch in S2 is less than 0.0625, it will be considered as a negative sample. However, if the IoA between it and one of the patches in S2 is more than 0.0625, it will be ignored.

3.4.2. Data augmented

We augment the data in order to improve the robustness of the algorithm and solve the problem that mitosis detection is insensitive to information such as position and angle. Moreover, we can improve the generalization ability of the model by increasing the amount of data without changing the image category.

Our positive samples are extracted from a large map with some mitotic cells. We do not rotate or translate each patch simply in data augmentation of positive samples. There will be blank areas that will lose information in the process of data augmentation and there is no white space for the patch of the test set. So, the general data augmentation method will affect the final result adversely. To solve this problem, we perform data augmentation operations in the original map and each patch contains complete background information. In this article, the maximum offset of translation is 16. We select five translation lengths less than 16 randomly. The offset should not be too large to ensure that the mitotic cells are in the middle of the patch. Since there are various angles of mitotic cells in the HPFs, we can augment positive samples by rotation. We rotate all patches six times randomly to further increase the amount of data after translation, .

4. Experiments

4.1. Dataset

We use MICCAI TUPAC Challenge dataset to evaluate our method. This dataset consists of 73 breast cancer cases from 3 different pathology centers. The first 23 cases are the dataset that was previously released as part of the AMIDA13 challenge, collected from the Department of Pathology at the University Medical Center in Utrecht. The remaining 50 cases are from two different pathology centers in The Netherlands (cases 24-48 are from one center and cases 48-73 are from another center). Each case is represented with some image regions with area of 2 mm^2. We have 656 2000×2000 pixel images in total. These images are extracted from whole slide images, which has been produced with the Leica SCN400 at a 40x magnification leading to a spatial resolution of 0.25μm and a 5657×5657 RGB image. The annotated mitotic figures are the consensus of at least two pathologists.

4.2. Training

4.2.1. Data Preprocessing

In the first step, the MITCAI TUPAC Challenge dataset provided a total of 656 2000×2000 pixel images as previously mentioned. Because the dataset contains cases from different laboratories, the staining results are different. The pictures under different staining conditions will bring additional interference to the network. For example, many false positives can

arise when the histopathology slide is overstained. Therefore, we will first use the method in Macenko et al. (2009) for staining unmixing and normalization.

In the second step, we cut the patches with the size of 128×128 in 2000×2000 pixel images based on the coordinates of the center point of the mitotic cell as described in Data Process in Our Method. We can get 1522 positive samples. To increase the robustness of the training model and the number of positive samples, the data is augmented with translation and rotation operation in the original image. There are 65082 positive samples after the data augmented.

In the third step, we cut the patches with the size of 128×128 according to the results of the second step and Eqs. 8. A box with the size of 128×128 traverses the full image from the upper left corner with a stride of 64 in each image. If the IoA of the patch is smaller than 0.0625, we set it as a negative sample. We obtain a total of 914579 negative samples after this operation.

In the fourth step, to improve the credibility of the experimental results and verify the robustness of the model, we use the 5-fold cross-validation method. The same patient's patches will be placed in the same set to ensure the validity of cross-validation.

4.2.2. Screening Network Setting

We train by Adam optimizer. The initial learning rate is set to 0.1, 0.01 after epoch 2, 0.001 after epoch 4, then 0.0001 after epoch 6 and 0.00003 after epoch 8. We use the Xvarier method to initialize all variables and we train for total 10 epochs. Our batch size is 192. We use moving average decay of 0.9, batch normalization moving decay of 0.99 and regularize decay of 2e-07. In the loss function as shown in Eqs. 6, α is 5 and γ is 2. By experiments, we set the threshold to 0.6 finally. Because of the large amount of parameters in the feature fusion part, we use the dropout layer in this part to avoid the over-fitting of the model and improve the generalization ability of the model. The dropout layer retains 50% of the parameters during training. Because the focal loss ignores many easy negative samples, the loss of the negative sample is reduced. The ratio of positive samples to negative samples fed to the refined network is 1:3 to ensure the accuracy of model prediction. To ensure the recall rate of positive samples is nearly 100% and screen out simple negative samples simultaneously, we set the threshold to 0.2. The result of the use of screening network on training dataset is in Table. 2.

4.2.3. Refined Network Setting

The input data for refined network consists of three parts: the negative samples set, the positive samples set and the false positive samples set. According to the previous description of hard example mining, the sampling ratio of the three sets is 1:1:3. We use the same parameters to initialize the network in the refined network part. Differently, the initial learning rate is set to 0.1, 0.01 after epoch 2.5, 0.001 after epoch 6, then 0.0001 after epoch 9.5 and 0.00003 after epoch 13. We train 15 epochs.

4.3. Testing

We perform a 5-fold cross-validation. First, we feed each patch of the validation dataset into the screening network and the refined network. If the probability of positive sample is

Table 2: The result of the use of screening network processing training dataset. FP, TP, FN represent the number of false positive samples, the number of true positive samples and the number of negative samples respectively. Recall(P/N) is the recall rate of positive samples/the recall rate of negative samples.

Fold	FP	TP	FN	Recall(P/N)	Precision	F1 score
1	83719	55056	377	0.9932/0.8792	0.3967	0.5669
2	98568	54254	210	0.9961/0.8724	0.3550	0.5234
3	84929	47227	325	0.9932/0.8758	0.3574	0.5256
4	86317	52542	376	0.9929/0.8649	0.3784	0.5479
5	104490	49532	213	0.9957/0.8659	0.3216	0.4862
AVE.	-	-	-	0.9942/0.8716		

greater than 0.80 in the screening network, and greater than 0.85 in the refined network, we will consider this patch as a positive sample. Finally, we use non-maximum suppression(NMS) to reduce the appearance of false positive. The suppressed result is our final result, as shown in Table. 3.

Table 3: The final result of 5-fold cross-validation. The AVE. means the average result of 5 folds. FP, TP, FN represent the number of false positive samples, the number of true positive samples and the number of negative samples respectively.

Fold	FP	TP	FN	Recall(P/N)	Precision	F1 score
1	23	153	76	0.6681/0.9999	0.8693	0.7556
2	10	102	150	0.4048/0.9999	0.9107	0.5604
3	21	209	208	0.5012/1.0000	0.9087	0.6461
4	4	140	149	0.4844/1.0000	0.9700	0.6467
5	79	273	92	0.7479/0.9994	0.7756	0.7615
Total	137	877	675	-/-	-	-
AVE.	-	-	-	0.5651/0.9999	0.8649	0.6836

4.4. Results analysis

4.4.1. EXPERIMENTAL RESULT

As shown in Table. 2, our network has successfully conducted the preliminary screening of data. The recall rate of the positive sample is greater than 99%, which indicates that the screening network almost selects all the positive samples and screens out a large number of negative samples. Refined network has a good ability to identify positive and negative samples. The result on the validation set is shown in Table. 3. Compared with the first place in TUPAC16 (Paeng et al., 2017), our experimental result reaches the most advanced level.

4.4.2. Experimental Contrast

In order to verify the effectiveness of semi-focal loss, hard example mining and cascaded network, our comparison experiments are shown in Table. 4. We experiment with the same data set and validation set of one fold.

Table 4: The result of comparison experiments with the same data set and validation set of one flod. CN means cascaded network. HEM means hard example mining. SFL means semi-focal loss while FL means ordinary focal loss. SM means softmax.

Method	Recall(P/N)	Precision	F1 score
CHS-NET(CN+HEM+SFL)	0.6681/0.9999	0.8693	0.7556
CN+HEM+FL	0.6419/0.9998	0.8258	0.7224
CN+HEM+SM	0.7579/0.9989	0.5585	0.6431
HEM+SFL	0.7816/0.9997	0.7217	0.7505
SFL	0.8471/0.9930	0.3012	0.4444

Through the above results, we can find out that cascaded network is able to screen out more negative samples and improve accuracy. Using hard example mining can obviously improve the F1 score compared with common classification network(SFL). Using the semi-focal loss is better than using focal loss alone, and the F1 score increase 3%. We can conclude that the cascaded network, the hard example mining and the semi-focal loss all improve the feature extraction ability and the classification ability of the model effectively.

The first place in TUPAC16 only uses the method of hard example mining, they obtain 0.652 F1-score on the test. It shows that our method can achieve state-of-the-art result by solving the problem of class imbalance. It also proves that the use of cascade network, hard example mining and semi-focal loss are effective.

5. Conclusion

The automatic mitosis detection of breast cancer histological images can help improve the efficiency and reliability of breast cancer screening and evaluation. The mitosis detection task is mainly faced with the problem of class imbalance. In this paper, we propose a new network called CHS-NET for mitosis detection. The screening network can detect 87% of negative samples and select false positives to increase the proportion of false positives in the dataset. We achieve 0.68 F1-score through a five-fold cross validation. This score is state-of-the-art compared with the best model in TUPAC16. It is proved by experiment that multi-level feature fusion can improve the accuracy of the model. Semi-focal loss can improve accuracy by 3%. It indicates that there are some inaccurate labels in the mitosis detection dataset of TUPAC16. From our work, researchers should pay attention to introduce some methods to solve the problem of inaccurate labels in the field of medical artificial intelligence.

References

Shadi Albarqouni, Christoph Baur, Felix Achilles, Vasileios Belagiannis, Stefanie Demirci, and Nassir Navab. Aggnet: Deep learning from crowds for mitosis detection in breast cancer histology images. *IEEE Transactions on Medical Imaging*, 35(5):1313–1321, 2016.

J. P. Baak, P. J. van Diest, F. J. Voorhorst, der Wall E Van, L. V. Beex, J. B. Vermorken, and E. A. Janssen. Prospective multicenter validation of the independent prognostic value of the mitotic activity index in lymph node-negative breast cancer patients younger than 55 years. *Journal of Clinical Oncology Official Journal of the American Society of Clinical Oncology*, 23(25):5993–6001, 2005.

Aayush Bansal, Xinlei Chen, Bryan Russell, Abhinav Gupta, and Deva Ramanan. Pixelnet: Representation of the pixels, by the pixels, and for the pixels. 2017.

Neslihan Bayramoglu and Janne Heikkilä. *Transfer Learning for Cell Nuclei Classification in Histopathology Images*. Springer International Publishing, 2016.

Hao Chen, Qi Dou, Xi Wang, Jing Qin, and Pheng Ann Heng. Mitosis detection in breast cancer histology images via deep cascaded networks. In *Thirtieth AAAI Conference on Artificial Intelligence*, pages 1160–1166, 2016.

D. C. Cireşan, A Giusti, L. M. Gambardella, and J Schmidhuber. Mitosis detection in breast cancer histology images with deep neural networks. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 411–8, 2013.

P J Van Diest, E Van Der Wall, and J P A Baak. Prognostic value of proliferation in invasive breast cancer: a review. *Journal of Clinical Pathology*, 57(7):675–81, 2004.

C. W. Elston and I. O. Ellis. 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(3a):403–410, 1991.

Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. pages 770–778, 2015.

Gao Huang, Zhuang Liu, Laurens Van Der Maaten, and Kilian Q. Weinberger. Densely connected convolutional networks. 2016.

Humayun Irshad. Automated mitosis detection in histopathology using morphological and multi-channel statistics features. *Journal of Pathology Informatics*, 4(1):10, 2013.

Tsung Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, and Piotr Dollar. Focal loss for dense object detection. pages 2999–3007, 2017.

Marc Macenko, Marc Niethammer, J. S. Marron, David Borland, John T. Woosley, Xiaojun Guan, Charles Schmitt, and Nancy E. Thomas. A method for normalizing histology slides for quantitative analysis. In *IEEE International Conference on Symposium on Biomedical Imaging: From Nano To Macro*, pages 1107–1110, 2009.

Kyunghyun Paeng, Sangheum Hwang, Sunggyun Park, and Minsoo Kim. A unified framework for tumor proliferation score prediction in breast histopathology. pages 231–239, 2017.

Florian Schroff, Dmitry Kalenichenko, and James Philbin. Facenet: A unified embedding for face recognition and clustering. In *Computer Vision and Pattern Recognition*, pages 815–823, 2015.

Abhinav Shrivastava, Abhinav Gupta, and Ross Girshick. Training region-based object detectors with online hard example mining. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 761–769, 2016.

K Sirinukunwattana, S Raza, Y. W. Tsang, D Snead, I Cree, and N Rajpoot. Locality sensitive deep learning for detection and classification of nuclei in routine colon cancer histology images. *IEEE Transactions on Medical Imaging*, 35(5):1196–1206, 2016.

Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. Going deeper with convolutions. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–9, 2015.

Mitko Veta, Paul J. Van Diest, Stefan M. Willems, Haibo Wang, Anant Madabhushi, Angel Cruz-Roa, Fabio Gonzalez, Anders B. L. Larsen, Jacob S. Vestergaard, and Anders B. Dahl. Assessment of algorithms for mitosis detection in breast cancer histopathology images. *Medical Image Analysis*, 20(1):237–248, 2015.

J. P Vink, M. B Van Leeuwen, C. H. M Van Deurzen, and G De Haan. Efficient nucleus detector in histopathology images. *Journal of Microscopy*, 249(2):124–135, 2013.

Erwan Zerhouni, Dvid Lnyi, Matheus Viana, and Maria Gabrani. Wide residual networks for mitosis detection. In *IEEE International Symposium on Biomedical Imaging*, pages 924–928, 2017.