Weakly Supervised Mitosis Detection in Breast Histopathology Images using Concentric Loss

Chao Li[a], Xinggang Wang[a], Wenyu Liu[a,*], Longin Jan Latecki[b], Bo Wang[c], Junzhou Huang[d]

[a] *School of Electronics Information and Communications, Huazhong University of Science and Technology, Wuhan, P.R. China.*
[b] *CIS Dept., Temple University, Philadelphia, PA 19122, USA.*
[c] *Department of Computer Sciences, Stanford University, USA.*
[d] *Tencent AI Lab, P.R. China.*

Abstract

Developing new deep learning methods for medical image analysis is a prevalent research topic in machine learning. In this paper, we propose a deep learning scheme with a novel loss function for weakly supervised breast cancer diagnosis. According to the Nottingham Grading System, mitotic count plays an important role in breast cancer diagnosis and grading. To determine the cancer grade, pathologists usually need to manually count mitosis from a great deal of histopathology images, which is a very tedious and time-consuming task. This paper proposes an automatic method for detecting mitosis. We regard the mitosis detection task as a semantic segmentation problem and use a deep fully convolutional network to address it. Different from conventional training data used in semantic segmentation system, the training label of mitosis data is usually in the format of centroid pixel, rather than all the pixels belonging to a mitosis. The centroid label is a kind of weak label, which is much easier to annotate and can save the effort of pathologists a lot. However, technically this weak label is not sufficient for training a mitosis segmentation model. To tackle this problem, we expand the single-pixel label to a novel label with concentric

*Corresponding author.
Email addresses: chaol@hust.edu.cn (Chao Li), xgwang@hust.edu.cn (Xinggang Wang), liuwy@hust.edu.cn (Wenyu Liu), latecki@temple.edu (Longin Jan Latecki), bowang87@stanford.edu (Bo Wang), joehhuang@tencent.com (Junzhou Huang)

circles, where the inside circle is a mitotic region and the ring around the inside circle is a "middle ground". During the training stage, we do not compute the loss of the ring region because it may have the presence of both mitotic and non-mitotic pixels. This new loss termed as "concentric loss" is able to make the semantic segmentation network be trained with the weakly annotated mitosis data. On the generated segmentation map from the segmentation model, we filter out low confidence and obtain mitotic cells. On the challenging ICPR 2014 MITOSIS dataset and AMIDA13 dataset, we achieve a 0.562 F-score and 0.673 F-score respectively, outperforming all previous approaches significantly. On the latest TUPAC16 dataset, we obtain a F-score of 0.669, which is also the state-of-the-art result. The excellent results quantitatively demonstrate the effectiveness of the proposed mitosis segmentation network with the concentric loss. All of our code has been made publicly available at `https://github.com/ChaoLi977/SegMitos_mitosis_detection`.

Keywords: Breast cancer grading, Fully convolutional network, Mitosis detection, Weakly supervised learning

1. Introduction

The most widely used invasive breast cancer grading system is the Nottingham Grading System (Elston & Ellis, 1991), which consists of three components: nuclear pleomorphism, tubule formation and mitotic count. Among them, mitotic count is the most important one since the propagation of cancer is mainly governed by cell division. Generally, the mitosis figures are marked manually by pathologists on the Hematoxylin and Eosin (H&E) stained slides. Counting mitosis manually is very time-consuming and subjective, thus it is extremely useful to develop an automatic detection method, which is capable of making this process more efficient and alleviating the bias of results from different pathologists.

However, detecting mitosis from H&E stained High Power Fields (HPFs) faces challenges in the following aspects: (1) There exist four phases (prophase,

metaphase, anaphase and telophase) during the development of mitosis, moreover, each phase among these four phases has very diverse shapes and texture configurations. In addition, a nucleus in the telophase splits into two distinct parts, but it should still be counted as one cell. Fig. 1(a) shows some mitotic cells with a diversity of appearances. (2) Many non-mitotic cells are very similar to mitotic cells in appearance, such as apoptotic cells and some dense non-mitotic nuclei. They all appear as dark blobs in H&E stained histopathology images. For instance, the last three samples in Fig. 1(b) show non-mitotic cells that have similar appearances with mitosis. (3) The processing of slide acquisition may introduce artifacts and unwanted objects, which further complicates this detection problem. Moreover, different conditions in preparation of tissues may also result in diversities in the appearances of the histopathology figures, which makes this task more challenging.

In recent years, many research works have been devoted to automatic mitosis detection. Most of these methods consist of two components: (1) A coarse method to locate the candidates of mitotic cells; (2) a more sophisticated classifier to classify the candidate patches produced from the first step. In this field, the most common method to locate candidate mitoses is the LoG (Laplacian of Gaussian) response on blue ratio images, which aims at finding darker regions. Then hand-crafted features or convolutional features are extracted from the candidate patches and fed into the following classifiers. The frequently used classifiers include SVM (Cortes & Vapnik, 1995), Random Forest (Breiman, 2001), AdaBoost (Freund & Schapire, 1995), deep convolutional neural network (LeCun et al., 1998), etc. Though IDSIA (Cireşan et al., 2013) directly uses a slide-window manner in the detection stage, the vast amounts of windows make it significantly time-consuming.

Unlike these methods, we propose a method that solves the mitosis detection task by the virtue of a semantic segmentation model. We predict the category of each pixel and then directly infer the locations of mitotic cells from the segmentation map.

Our method, called SegMitos, is illustrated in Fig. 2. We train a segmen-

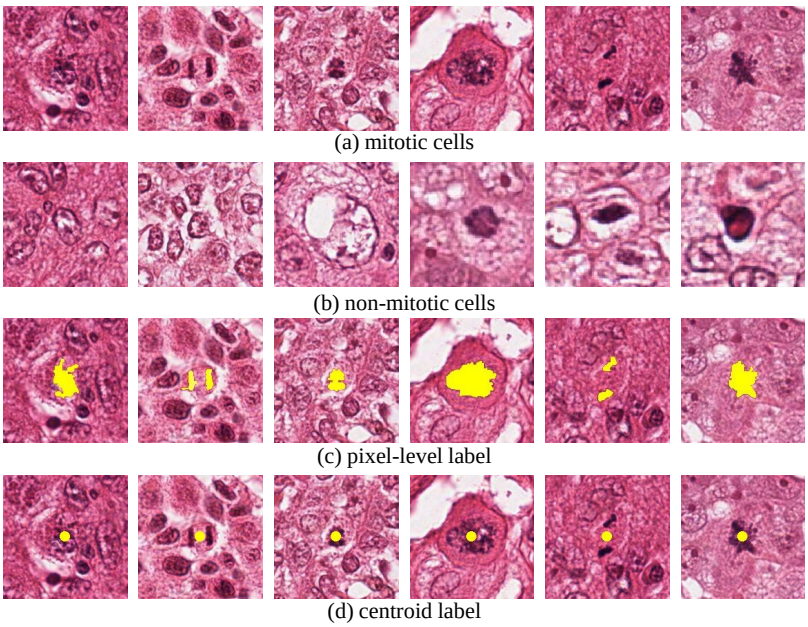


Figure 1: Examples of different types of cells and labels. (a) shows mitotic cells. (b) shows non-mitotic cells where the last three samples have similar appearances with mitotic cells. (c) shows the pixel-level label of the mitoses in (a). (d) shows the centroid label of the mitoses in (a). The single annotated pixel in (d) is represented by a large dot for better view.

tation network based on fully convolutional network (FCN) (Long et al., 2015) with the mitosis data. The SegMitos model produces a segmentation map where each pixel represents its confidence of belonging to the mitosis class. We process the response map with a Gaussian filter in order to reduce image noise. Then an adaptive binarization process is applied to yield segmentation blobs. For these blobs, we calculate their areas and mean confidence scores. Finally, we obtain detection results by using a filtering mechanism based on these two features. The SegMitos model is applied in an end-to-end (image-to-image) fashion without using sliding window method, hence the detection is very fast and efficient.

In mitosis benchmark datasets, there are two types of annotations. The first one is pixel-wise that precisely annotates all pixels of every mitotic cell, as shown in Fig. 1(c). The second one is the centroid label as shown in Fig. 1(d) which only marks a single pixel in the central zone of mitotic cell. To train a segmentation model, the pixel-wise label is suitable while the centroid label is not. Therefore, we propose a novel concentric label based on the weak centroid annotation. Firstly, we use a circle which centered at the annotated pixel as an estimated mitotic region. Then considering that there may be some mitotic pixels outside the circle, we encompass the mitotic circle using a larger circle. We take the ring between the two circles as a neutral region. We ignore the computation of the loss within the "middle ground" so that this region makes no contribution to the parameter learning of the network. The mitotic circle and neutral ring constitute a concentric label, and the tailored loss computation method forms a concentric loss. By means of the proposed concentric label and loss, we can train a segmentation model in a weakly supervised way.

To evaluate the proposed method, we conduct experiments on four mitosis detection datasets. On the ICPR 2014 MITOSIS dataset (MITOS-ATYPIA-14, 2014) and AMIDA13 dataset (Veta et al., 2015), we achieve better performance than the state-of-the-art methods with remarkable improvements. On the latest TUPAC16 (Veta et al., 2018) dataset, our method also produces the best performance.

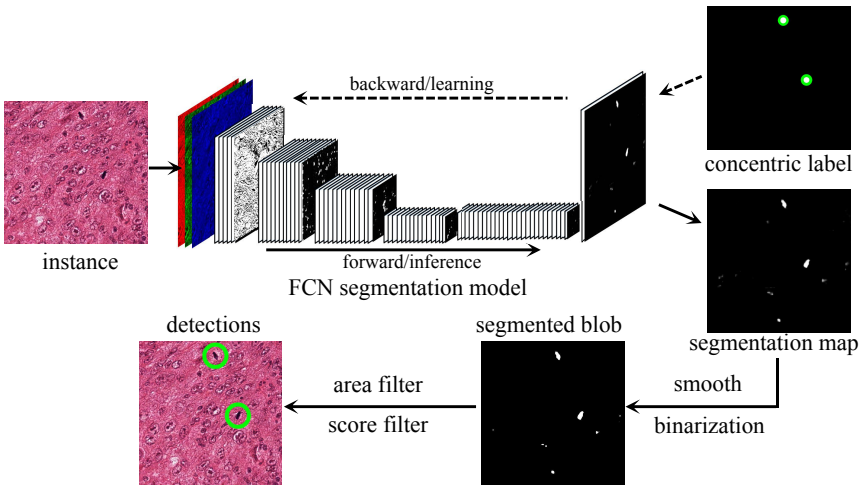


Figure 2: The SegMitos model trained with concentric label for mitosis detection. In the label map, the small white circle denotes the estimated mitotic region and the green ring denotes the ignored region when computing loss. Segmentation network yields a pixel-wise segmentation map. Then a smooth and a binary processing are applied to the segmentation map, aiming to yield some segmented blobs. Finally, we apply a morphological filtering step based on the confidence score and area of the segmented objects. The green circles in the bottom left corner are true positives.

In this paper, we mainly have three contributions. (1) We propose a novel concentric label as well as a concentric loss, making it possible to train a dense prediction model with weak annotation. To our knowledge, this is the first research that formulates a concentric loss function to train a semantic segmentation network for weakly supervised mitosis detection. (2) We design an automatic and practical system for detecting mitosis in histopathology images. We validate our method on almost all mitosis detection benchmarks and achieve state-of-the-art performances on all these datasets. (3) We deploy a semantic segmentation model to solve the mitosis detection problem. We abandon the common two-step strategy which is composed of segmentation and classification. Though the filtering of the segmented candidates in our method can be regarded as a classification procedure to a certain extent, our classifier is very straightforward and merely utilizes the output of segmentation model as features. Our method mainly focuses on the segmentation model while other methods focus on the patch-based classifiers.

2. Related work

Recently, many methods for automatic mitosis detection from histopathology slides images have been proposed. Some methods (Sommer et al., 2012; Veta et al., 2013; Irshad et al., 2013; Tek et al., 2013; Paul & Mukherjee, 2015) utilize a variety of hand-crafted features, including morphological and statistical features. The most commonly used morphological features are area, perimeter, eccentricity, major axis length, minor axis length, equivalent diameter, etc. While the frequently used statistical features include the mean, median, variance of various color channels, color histogram features, color ratios, etc. Researchers try to use these features to reflect the appearance characteristics of mitosis. However, since the mitotic cells have varieties of appearances, it is hard to manually design features to distinguish mitosis from normal cells very accurately.

Convolution neural networks (CNN) based methods have revolutionized the field of computer vision. They have achieved many state-of-the-art results on a

variety of tasks including image classification (Krizhevsky et al., 2012; Simonyan & Zisserman, 2014; Szegedy et al., 2015; He et al., 2016; Tang et al., 2016, 2017; Wang et al., 2018), object detection (Girshick et al., 2014; Girshick, 2015; Ren et al., 2015; Li et al., 2017; Tang et al., 2018), and image segmentation (Long et al., 2015; Xie & Tu, 2015; Shen et al., 2017b). Recently, many deep learning based methods (Shen et al., 2017a; Zhou et al., 2017; Wang et al., 2017; Yin et al., 2018) have been applied to medical image analysis tasks and achieve promising results. Shen et al. (Shen et al., 2017a) take advantage of multi-stage multi-recursive-input fully convolutional networks(M^2FCN) to detect neuronal boundary. In their proposed framework, the multiple side outputs from the lower stage are fed into the next stage, providing multi-scale contextual information for the consecutive learning. To improve the segmentation accuracy of pancreas, Zhou et al. (Zhou et al., 2017) apply a fixed-point model using a predicted segmentation mask to shrink the input region. The smaller input region leads to more accurate segmentation for this small organ.

Some CNN based approaches have also been proposed for addressing the mitosis detection problem. IDSIA (Cireşan et al., 2013) uses a deep convolutional neural network as a pixel classifier. The classifier is trained to distinguish mitotic patches from normal patches. In the detection stage, the trained classifier is applied to breast histopathology images with a sliding window manner, which is computationally demanding. Though the IDSIA won the 2012 mitosis detection contest (Roux et al., 2013) and AMIDA13 challenge (Veta et al., 2015), the poor efficiency prohibits its application in clinical practice. Wang et al. (Wang et al., 2014) produce candidate patches and then utilize classifiers to classify these patches. They take advantage of hand-crafted features and convolutional features to train classifiers separately. To improve the accuracy, they also use both types of features to train a more accurate classifier. CasNN (Chen et al., 2016a) uses two networks to build a cascade detection system. The first network is a coarse retrieval network that locates candidates of mitosis, and the second network is a fine discrimination model that classifies these candidate patches. Lunit (Paeng et al., 2017) trains a deep classification CNN with image patches.

In order to reduce the false positive rate, it uses a two step training procedure. DeepMitosis (Li et al., 2018) applies a proposal-based deep detection network to the mitosis detection task and utilizes a patch-based deep verification network to improve the accuracy. In our work, instead of using the convolutional network as a patch classifier, a patch feature extractor or an object detector, we deploy the network to perform semantic segmentation.

Many methods firstly segment the images to seek candidates of mitosis, and then classify these candidates to produce the final detections. Irshad (Irshad et al., 2013) transforms the RGB images to blue ratio images and then computes the LoG response on them. Candidates are produced through morphological processing, and then be classified by a decision tree classifier. Wang et. al (Wang et al., 2014) utilize a similar candidate segmentation method as Irshad (Irshad et al., 2013), but apply convolutional neural network and Random Forest (Breiman, 2001) as classifiers. Paul and Mukherjee (Paul & Mukherjee, 2015) segment cells using the maximization of relative-entropy between the cells and the background, and then classify the segmented cells by the Random Forest classifier.

Unlike these approaches, we use a fine segmentation model to produce pixel-level predictions and then directly obtain detection results from the segmentation probability map. Our method mainly focuses on the segmentation part. Though we also classify the segmented candidates, the classification stage in our method is very straight-forward and easy. On the contrary, others' methods usually focus on the classification of mitosis candidates. They use complex model, such as SVM, Random Forest or CNN, to judge the candidates. The segmentation stages in their system are generally rough and not significant.

3. Approach

In this section, we describe the proposed SegMitos method for mitosis detection and the concentric loss for training the SegMitos model.

3.1. Mitosis segmentation based on fully convolutional network

3.1.1. Fully convolutional network

Our SegMitos model is based on a fully convolutional network designed for semantic segmentation on natural images. Fully convolutional network (Long et al., 2015) adapts all fully connected layers in conventional classification network to convolutional layers. It can take an input of arbitrary size and produce an output of the corresponding size by performing upsampling in-network. The output of FCN reserves the spatial structure of the original image. Thus FCN is very suitable to perform dense prediction tasks such as semantic segmentation. FCN has an end-to-end (image-to-image) learning architecture. It adopts a whole image training strategy, which is more efficient than the patch-wise training.

The stride in the final convolutional layer of FCN is 32 pixels. The relatively large stride makes the prediction map very coarse and prone to losing details. To address this problem, FCN exploits the prediction output from the lower layer to fuse with the original prediction map. The lower layer has a smaller stride and receptive field size compared with the high layer, so its prediction is finer and more related to the local structure. Combining different prediction layers can exploit complementary information in them to yield a better segmentation result. As shown in Fig. 3, fusing the prediction of the base FCN-32s with the prediction from *pool*4 layer yields FCN-16s model, which has a 16-pixel stride. Further combining the prediction of FCN-16s with the prediction from *pool*3 layer yields the network named FCN-8s whose stride is 8 pixels.

3.1.2. SegMitos for mitosis segmentation

The SegMitos model derived from FCN is tailored for performing segmentation on breast histopathology images. Since this task is a two-category (mitotic cell or background) detection problem, we set the channel number of prediction layers in the SegMitos model to two. As the prediction result is on pixel level, we can obtain the fine shape of mitotic cell. Our detection result is finer than

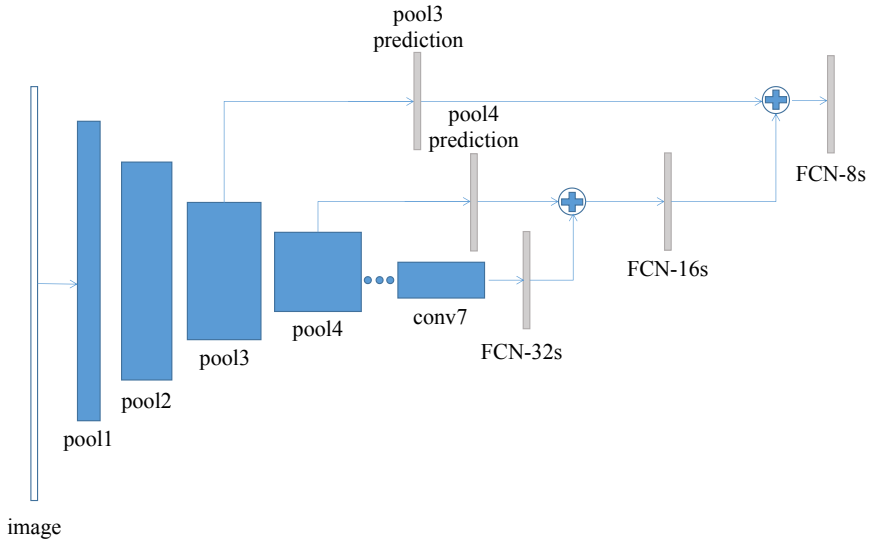
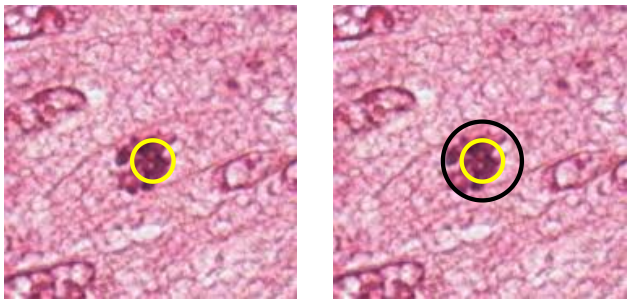


Figure 3: FCN framework with different strides. The $32s$ in FCN-32s denotes the stride pixels in the segmentation map before upsampling. So as the FCN-16s and the FCN-8s.

those of the common detection methods, which only give out results in bounding box format.

3.1.3. Transfer learning across domains

The training of deep CNN usually requires a great deal of data to optimize the enormous number of network parameters. However, in the medical image domain, the data is usually limited and the expert annotation is expensive. Thus, training a deep learning model from scratch is usually not practical in medical image analysis. Fine-tuning is a widely used method for transfer learning in neural network training. Previous work (Tajbakhsh et al., 2016) has validated the superiority of pre-trained CNN with adequate fine-tuning, indicating the efficacy of transfer learning from natural image domain to medical image domain. Herein, we fine-tune the FCN model which has been pre-trained on the Pascal VOC image segmentation task (Everingham et al., 2010). We initialize our networks with parameters from the pre-trained FCN and fine-tune all layers through back-propagation on mitosis data.



(a) circle label (b) concentric label

Figure 4: Visualization of the proposed mitosis label. (a) shows the circle label that intends to involve mitosis pixels. (b) shows the concentric label of which the ring region is the middle ground to comprise the pixels that are hard to estimate. The pixels inside the yellow circle are almost mitotic while the ring region between the two circles includes both mitotic and non-mitotic pixels.

3.1.4. Data augmentation

The size of a full HPF image is usually very large that directly processing the full image will occupy too much GPU memory. Hence we crop patches from the original images. The side length of the patch is about 500 pixels. To produce more training samples, overlapping patches are sampled.

Due to the limited number of mitosis images, data augmentation is very critical to effectively train a deep model and prevent over-fitting. We perform some image transformations on histopathology images, including image rotation, translation, and mirroring. As the proportion of mitotic cells is usually low in histopathology images, the distribution of mitotic/non-mitotic pixels is heavily biased. To alleviate the data imbalance, we utilize various kinds of image transformation to produce much more variants for the patches that contain mitotic pixels than the negative patches.

3.2. Concentric label and loss

In this paper, we use four datasets to validate the proposed approach. The first dataset is the 2012 ICPR MITOSIS dataset (Roux et al., 2013), which provides pixel-wise labels as shown in Fig. 1(c). This type of annotation can

reflect the shape of a mitotic cell accurately, providing a strong supervision to train a semantic segmentation model. However, labeling every pixel of mitoses is very time-consuming for pathologists, so the pixel-wise annotation is not the most common label for mitosis data. Most datasets adopt the centroid label as shown in Fig. 1(d). The other three datasets (i.e., the 2014 ICPR MITOSIS dataset (MITOS-ATYPIA-14, 2014), the AMIDA13 dataset (Veta et al., 2015) and the TUPAC16 dataset (Veta et al., 2018)) used in our experiments adopt the centroid label. This type of annotation only marks a central pixel of a mitotic cell, which is much easier and faster for pathologists to annotate. However, the centroid label does not provide an accurate shape of mitotic cells and hence cannot be used to train dense prediction model directly. To tackle this problem, we design two solutions for the centroid label.

The first one is a circle centered at the original centroid pixel, as shown in Fig. 4(a). We regard the circular region as a mitotic cell. Based on a statistic analysis in 2012 MITOSIS training set, the average area of mitotic cells is about 590 pixels. Considering that the shape of a mitotic cell is irregular and unknown, we make the circle small so that the pixels within the circle incline to be positive. Specifically, we try two different configurations in circle label: the circle label with 12-pixels radius and the circle label with 15-pixels radius. More details will be described in Section 4.3.

However, there may still exist some mitotic pixels outside the circle. Roughly regarding all pixels outside the circle as negative pixels would impede the optimization of segmentation networks. Thus, we add a ring region surrounding the small circle as the "middle ground". Pixels in this ring region have no strong possibility to be positive or negative so we regard them as neutral samples. Specifically, this new label is depicted as a format of two concentric circles as shown in Fig. 4(b). The small yellow circle encircles the mitotic pixels as the circle label, while an extra black circle is added to generate a ring region. Since the shape of a mitotic cell is irregular, the ring region usually contains both mitotic pixels and non-mitotic pixels. It is hard to estimate correct labels for these ambiguous pixels, so we ignore them during the training stage. It can

be noticed that the ring region in Fig. 4(b) indeed contains mitotic pixels and background simultaneously, so it is reasonable to assign neutral labels to them.

Since mitotic cells have varieties of shapes and areas, assigning a fixed concentric circle label to all cells is not appropriate. Thus, we propose a random radii configuration for concentric circle label. Here we define the radius of the small circle as r and the radius of the large circle as R. The r is randomly chosen from a uniform distribution between 10 and 17 pixels, and the R is randomly 1.5 times to 2.5 times the length of the r. We use multiple random concentric labels for each mitotic cell, which could better fit the true morphological shape of a cell.

We design a corresponding pixel-wise loss function named concentric loss for the concentric label as following:

$$\mathbf{L} = -\sum_{n=1}^{N}(\sum_{x\in C_1}\log P(1|x;W) + \sum_{x\in B}\log P(0|x;W)) + \lambda\sum_{n=1}^{N}\|W\|_2^2 \qquad (1)$$

where x is the pixel, W is the weight of network. $P(1|x;W)$ is the probability of x being mitotic pixel given the network weight W. N is the number of images, C_1 is the round mitotic region inside the small circle, B is the background region outside the big circle. The ring region between the C_1 and B does not contribute to the loss computation. The last part is a regularization term and the λ controls the balance. Our concentric loss is based on the Softmax loss. We optimize the parameters of SegMitos model on the histopathology images by minimizing the concentric loss function.

3.3. Inferring mitosis on segmentation map

After training the SegMitos model, we apply it to segment the breast histopathology images and produce pixel-wise predictions. Based on the segmentation probability map, we infer mitotic cells by using a heuristic method.

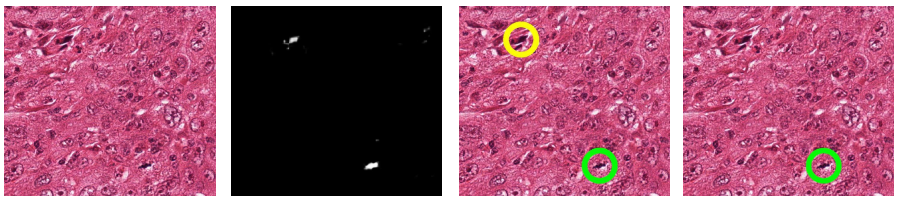
For the sake of GPU memory, we divide the test HPF images into small patches. The output of SegMitos is a segmentation map in which the score of a pixel denotes its confidence of belonging to the mitosis class. We stitch the segmentation maps of these patches to get the response map of the full

HPF. The prediction map is noisy due to the existence of ambiguous cells and noise. To smooth the map, we apply a Gaussian filter on the segmentation response map. The smoothing can eliminate some tiny noisy responses. Then we apply an image binarization on the response map to generate segmented blobs. Here we take an adaptive threshold in the binarization of images. The threshold is determined by Otsu's method (Otsu, 1975), which minimizes the intra-class variance of each class. Our threshold is determined holistically from the segmentation map of a full HPF, rather than in a patch-wise way.

We regard the segmented blobs as candidates for mitosis. By designing a filtering mechanism, we eliminate false detections and obtain the true mitotic cells. Specifically, we define two features on the segmented blob. The first one is the area of the blob, and the second one is the blob's mean confidence score. Intuitively, the response of mitosis in segmentation map should be large and bright while the response of normal cells should be small and dim. Based on this assumption, we filter the blobs according to the areas and mean scores. For a blob, if its mean score $\bar{S}$ is higher than a threshold $s1$ and meanwhile its area A is larger than a threshold $a1$, this blob will be kept as a detection, otherwise, it will be filtered out. This threshold mechanism can efficiently and effectively remove some false responses caused by easily-confused pixels, like the nuclei of some non-mitotic cells. Fig. 5 illustrates the visual effect of each step in our method. In this example, merely using the area feature to filter the segmented blobs would produce a false positive. Further taking advantage of the filter based on the confidence score can remove this false detection.

4. Experiments

In this section, we evaluate the performance of our approach on four widely used mitosis detection datasets and achieve state-of-the-art results on all of them. We also use our concentric label on a breast mass segmentation dataset to demonstrate that the proposed method can be extended to other segmentation tasks.



(a) original image (b) segmentation map (c) filter by area (d) filter by area and score

Figure 5: Visual effect of each step of our proposed method. (a) shows a sample region from a H&E slide image. (b) shows the visual result of segmentation probability map. (c) shows the detection results of merely using the area feature to filter blobs. The green circle denotes true positive while the yellow circle denotes false positive. Some tiny responses are screened by this step. (d) shows the final detection results of using area and mean score to filter.

4.1. Implementation

We use a TITAN X GPU with 12GB memory to run our experiments. We train our SegMitos model using the Caffe framework (Jia et al., 2014). The initial model is FCN-32s which is finetuned from VGG-16 net (Simonyan & Zisserman, 2014) on PASCAL segmentation dataset. This model is publicly available on the model zoo site of Caffe [1].

Model parameter. We follow the default training configuration of FCN. A small learning rate (1e-10) and a high momentum (0.99) are used. The weight decay is 0.0005 and the batch-size is 1. We first train the base version of model that has 32-pixels stride, and then we train the two finer versions of SegMitos model. The SegMitos-16s model is trained based on the SegMitos-32s model with a learning rate 1e-13. Finally, SegMitos-8s model is trained based on the SegMitos-16s with a learning rate 1e-14. The networks are optimized by stochastic gradient descent.

Measurement of performance. According to the criteria of the mitosis detection contests, a detection would be counted as a true positive when its centroid is within a defined distance from the centroid of a ground truth mitosis.

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

In 2012 MITOSIS dataset, the threshold of distance is 5 μm (20 pixels). In the other three datasets, this threshold of distance is 7.5 μm (30 pixels). The accuracy of detection results is evaluated using F-score, which is the harmonic mean of recall and precision, as shown below.

$$F_{score} = 2 \times recall \times precision/(recall + precision) \tag{2}$$

4.2. 2012 MITOSIS dataset

The 2012 ICPR MITOSIS dataset (Roux et al., 2013) is extracted from a set of five breast cancer biopsy slides. In each slide, the pathologists selected 10 HPFs at 40X magnification. Among the 50 HPFs, 35 HPFs are used for training and the remained 15 HPFs for testing. The number of mitotic cells in the training set and test set are 226 and 101, respectively. The spatial size of HPF is 2084×2084 pixels. The annotation of this dataset is strongly supervised that all pixels of a mitotic cell are labeled.

In this dataset, we rotate the negative patch in 4 angles and mirror it. While for a positive patch, we apply more data augmentations including image rotation in 16 angles, image translation in 9 directions and image mirroring.

4.2.1. Ablation experiments

We conduct some ablation experiments using the SegMitos-32s model to analyze the effect of each component in our framework. Results are reported in Table 1 and discussed in detail next. The experiments are run on the 2012 MITOSIS test set.

Impact of smoothing the map. Removing the Gaussian smooth operation of the response map results in a 0.4 percent loss in performance. This indicates that the smoothing can remove some tiny noisy response and improve the precision.

Impact of holistic binarization. In our method, we piece together the segmentation maps of the patches and then determine the binarization threshold

Table 1: Ablations for SegMitos-32s on 2012 MITOSIS test set.

smooth?		✓	✓	✓	✓	✓
holistic binarization?	✓		✓	✓	✓	✓
use area?	✓	✓			✓	✓
use score?	✓	✓		✓		✓
F-score(%)	78.78	77.61	66.15	75.24	78.39	**79.19**

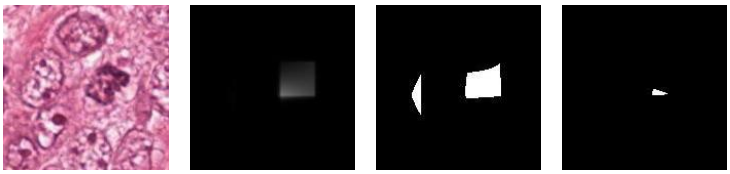


Figure 6: The comparison of patch-wise binarization and holistic binarization. (a) shows a histopathology image sample from a patch. (b) shows the segmentation map of (a) produced by SegMitos model. (c) shows the patch-wise binarization result. (d) shows the holistic binarization result. There exist no mitosis in this image sample.

by using the Otsu's method on the full HPF's segmentation map. To validate the superiority of holistic binarization, we conduct a controlled experiment to perform binary segmentation on patches individually. The thresholds of each patch are produced by applying Otsu's method individually. This patch-wise binary segmentation causes a 1.5 percent loss in performance. It is partly because in some patches that have no high response, the threshold would be low, and that would introduce more extra noisy blobs compared to the holistic threshold. In Fig. 6, we show the comparison of these two binarization ways. Due to the low response of the segmentation probability map of this patch, the patch-wise Otsu's binarization gets a too low threshold that two false segmented blobs are produced. In contrast, the holistic binarization can achieve an appropriate threshold by considering the response of full image, so it produces a much better binary result.

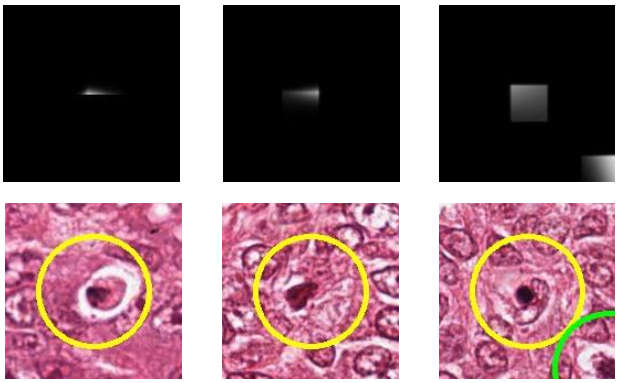


Figure 7: (a) and (b) show two false positive examples if we discard the area information. (c) shows one false positive if we do not use confidence score information. All these false detections can be screened out when we use both two features. The first row shows the SegMitos model's segmentation maps and the second row shows the detection results on original images. The yellow and green circles are false positives and true positives, respectively.

Impact of using area information. If we do not utilize the area information when filtering the candidates, the F-score would drop almost four percent. So the area of segmented blobs is very critical to distinguish the mitotic cells from normal cells. There exist some tiny blobs with relatively high mean response scores in the segmentation map. They are probably the nuclei of non-mitotic cells or mimic pixels. For instance, the mean scores of the two blobs in Fig. 7(a) and (b) are not low. If we only use the score information, it will produce two false positives. Taking advantage of their area information can effectively screen them out.

Impact of using score information. If we do not utilize the mean confidence score, the F-score drops 0.8 percent. We infer that the mean score of the blob is useful but not as important as the area information. In Fig. 7(c), the area of the cell's response on segmentation map is not small. Merely using the area information cannot filter it out.

We then analysis the change of performance with respect to the two features'

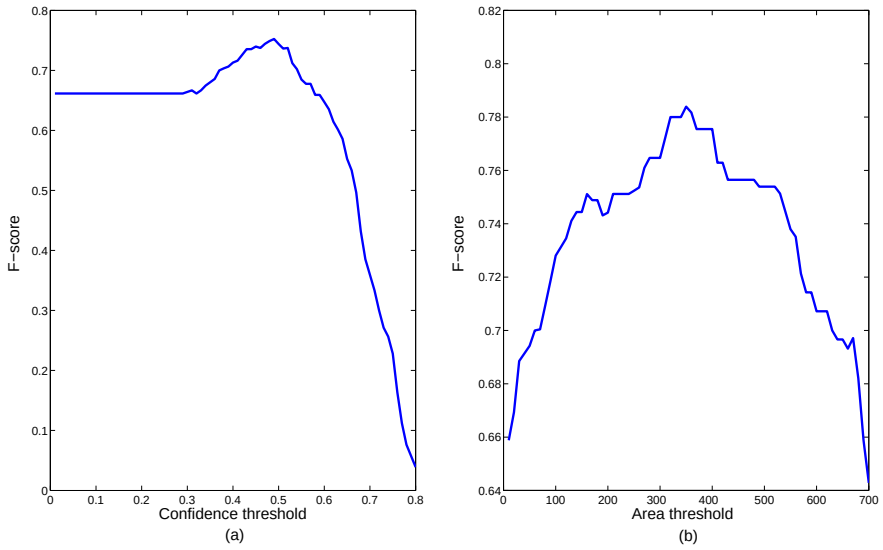


Figure 8: The performance of using only one feature to filter the candidates in 2012 MITOSIS test set. (a) shows the change of F-score with respect to the confidence threshold, when we only use the mean confidence to filter the candidates. (b) shows the change of F-score with respect to the area threshold, when we only use the area information to filter the candidates.

thresholds, respectively. In Fig. 8(a), we shows the performance when we only use the confidence score in the filtering stage. The F-score is relatively robust with respect to the confidence threshold. Fig. 8(b) shows the change of F-score with respect to the area threshold. It is noted that the F-score is not very sensitive to the area threshold. A wide range of area threshold (200-500 pixels) can achieve good performance. Moreover, if we directly regard all the segmented blobs as the final detections, without any filtering, we can still achieve a 0.662 F-score. This result indicates our segmentation model is very powerful.

We further evaluate the finer stride versions of SegMitos model. Table 2 contains the performance comparison of different stride versions of SegMitos model. It can be noticed that the SegMitos-16s model achieves an inferior performance compared with the 32s model, probably because the training of 16s model sinks into a local convergence. While the finest model SegMitos-8s actually achieves the best result among them.

Table 2: Performance of different SegMitos models on 2012 MITOSIS test set.

Model	Precision(%)	Recall(%)	F-value(%)
SegMitos-32s	81.25	77.23	79.19
SegMitos-16s	76.70	**78.22**	77.45
SegMitos-8s	**84.61**	76.24	**80.21**
SegMitos-random	81.31	73.27	77.08

We also train a SegMitos-32s model with the random concentric label. It achieves a 0.7708 F-score, a comparable performance to the models trained with pixel-level labels. Though we only utilize the centroid information of mitosis in the concentric label, we can still obtain a very accurate semantic segmentation network. The performance gaps between the models trained with strongly supervised labels and the one trained with our concentric label are not large, which demonstrates the validity of the proposed concentric loss. Here, we do not use the concentric label to train a 16s or 8s model, because we think the estimated concentric label is not very precise that it is hard to further improve the performance by training finer versions of model.

In Fig. 9 we show some detection results of our method on test histopathology images. It can be noted that the false positives have very similar appearance with mitotic cells so it is hard to filter them out. The undetected mitotic cells are sometimes not obvious enough or too small to be discovered.

Table 3 presents the results of SegMitos and other methods. The last four methods (IDSIA (Cireşan et al., 2013), IPAL (Irshad et al., 2013), SUTECH (Tashk et al., 2013) and NEC (Malon et al., 2013)) in Table 3 are participants of 2012 MITOSIS challenge, while all other methods appeared after the contest. Our SegMitos yields rank three of all methods. The excellent result qualitatively validates the effectiveness of deploying a semantic segmentation model to the mitosis detection application.

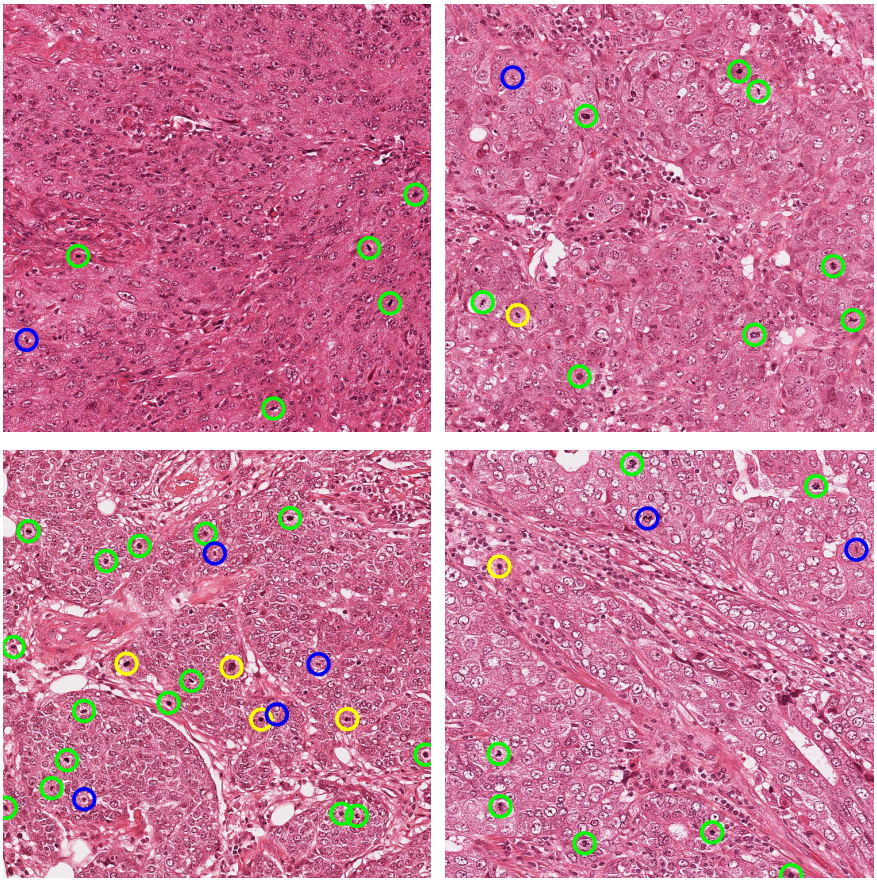


Figure 9: The detection samples of our SegMitos method on 2012 MITOSIS test set. The green, yellow and blue circles are true positives, false positives and false negatives, respectively.

Table 3: Performance of different methods on 2012 MITOSIS test set.

Method	Precision(%)	Recall(%)	F-value(%)
SegMitos-8s	84.61	76.24	80.21
SegMitos-random	81.31	73.27	77.08
DeepMitosis (Li et al., 2018)	85.4	**81.2**	**83.2**
RRF (Paul et al., 2015)	83.5	81.1	82.3
CasNN (Chen et al., 2016a)	80.4	77.2	78.8
HC+CNN (Wang et al., 2014)	84.0	65.0	73.45
IDSIA (Cireşan et al., 2013)	**88.6**	70.0	78.2
IPAL (Irshad et al., 2013)	69.81	74.0	71.84
SUTECH (Tashk et al., 2013)	70.0	72.0	70.94
NEC (Malon et al., 2013)	75.0	59.0	65.92

4.3. 2014 MITOSIS dataset

We then evaluate SegMitos on the 2014 ICPR MITOSIS dataset (MITOS-ATYPIA-14, 2014). This dataset has eleven folders in training data, where each folder contains the HPF images from one breast cancer biopsy slide. There are 1200 HPFs in training data, which include 749 mitotic cells. The spatial size of HPF in this dataset is 1539×1376 pixels. In contrast to the 2012 dataset, pathologists only annotate the centroid pixel for each mitotic cell in this dataset. There are totally 496 HPFs from five folders (slides) in the test set. The annotation of the test set is held out by organizers and the number of mitosis cells is unknown.

We select the first folder $A03$ from training data as the validation set, and the next seven folders ($A04, A05, A07, A10, A11, A12, A14$) as the training set. The training set contains 816 HPFs and 534 mitosis cells, while the validation set contains 96 HPFs and 135 mitotic cells. We crop patches sized 385×344 pixels from full HPF for the sake of memory.

Since the number of images in 2014 MITOSIS dataset is much larger than

that of 2012 dataset, we take fewer image transformations on images of 2014 dataset. Specifically, for a patch that contains mitotic pixels, we translate, rotate, and flip it to generate 144 variants. While for a negative patch, we do not deploy augmentation as the negative pixels are already adequate in this dataset.

The weights of SegMitos model are initialized from the model trained on 2012 MITOSIS dataset. In this dataset, we only use the 32-pixels stride architecture due to the coarse label. The parameters (area threshold and mean confidence threshold) of our filtering mechanism are optimized on the validation set.

Table 4 contains the detailed results of different SegMitos models on 2014 MITOSIS validation set. The SegMitos(2012) in the first row is the SegMitos-32s model trained on 2012 MITOSIS dataset, while all other models are fine-tuned on 2014 MITOSIS training set. To compare our proposed circle label and concentric label, we use these two kinds of annotations to train the segmentation model, respectively. For the circle label, we try two different configurations of radius. The SegMitos-r12 model is trained with a circle label whose radius is 12 pixels. Similarly, the SegMitos-r15 model uses a circle label whose radius is 15 pixels. For the concentric circle label, besides the random radii, we also try the configuration of fixed radii in this dataset. The SegMitos-r12R24 model utilizes a concentric label, whose small radius is 12 pixels and large radius is 24 pixels. The next model SegMitos-r15R30 is also trained with concentric label but with different radii of circles. The SegMitos-random model is trained with the random concentric circle label. As described in Section 3.2, the radius r of small circle in random concentric label is chosen from a uniform distribution between 10 and 17 pixels, and the ratio of R to r is 1.5 to 2.5.

The large performance gaps between the SegMitos(2012) model and the other models fine-tuned on 2014 MITOSIS images indicate the significance of fine-tuning. The models using concentric label outperform the models using circle label, which demonstrates the superiority of the proposed concentric label and loss. Fig. 10 shows some detection results produced by SegMitos-random model on 2014 MITOSIS validation set.

Table 4: Results on 2014 MITOSIS validation set.

Method	Precision(%)	Recall(%)	F-score(%)
SegMitos(2012)	42.56	61.48	50.30
SegMitos-r12	48.02	71.85	57.57
SegMitos-r15	40.14	**82.96**	54.10
SegMitos-r12R24	51.68	68.15	58.78
SegMitos-r15R30	49.53	78.51	**60.74**
SegMitos-random	**54.12**	68.15	60.32

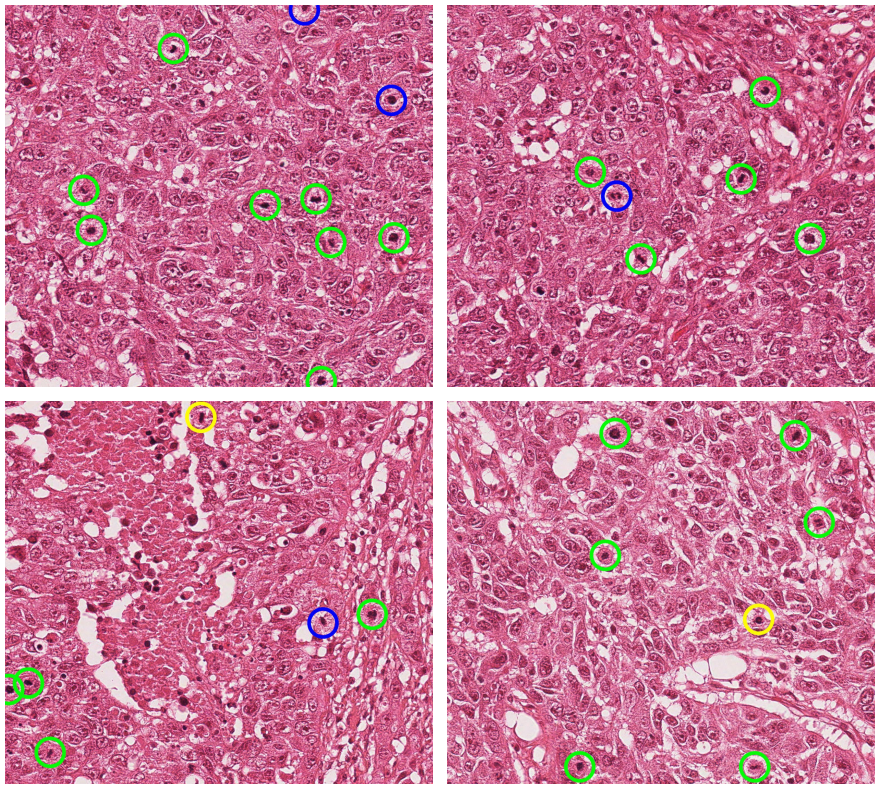


Figure 10: The detection samples of our SegMitos-random model on 2014 MITOSIS validation set. The green, yellow and blue circles are true positives, false positives and false negatives, respectively.

Table 5: Performance of different approaches on 2014 MITOSIS test set. "-" denotes the results which are not released.

Method	Precision(%)	Recall(%)	F-score(%)
STRASBOURG	-	-	2.4
YILDIZ	-	-	16.7
MINES-CURIE-INSERM	-	-	23.5
CUHK	44.8	30.0	35.6
DeepMitosis (Li et al., 2018)	43.1	44.3	43.7
CasNN(single) (Chen et al., 2016b)	41.1	47.8	44.2
CasNN(average) (Chen et al., 2016b)	46.0	50.7	48.2
SegMitos-r12	46.88	**51.72**	49.18
SegMitos-r12R24	62.25	46.31	53.11
SegMitos-r15R30	59.43	51.23	55.03
SegMitos-random	**63.75**	50.25	**56.20**

We run the fine-tuned SegMitos models on the 2014 MITOSIS test set and submit the detection result to the organizers of the challenge. The evaluated F-scores of our models and some other methods are shown in Table 5. CUHK, MINES-CURIE-INSERM, YILDIZ and STRASBOURG are the top four approaches that participate the 2014 ICPR MITOS-ATYPIA contest (MITOS-ATYPIA-14, 2014). The "single" version of CasNN (Chen et al., 2016b) utilizes a single classification network in its fine discrimination stage, while the "average" version fuses results of multiple classification networks. As we can see, our SegMitos methods outperform previous state-of-the-art methods. The markedly superior detection performance results from the powerful discriminative capacity of SegMitos model. The model trained with the concentric loss using the random radius, termed as SegMitos-random, achieves the best performance among all SegMitos models. It indicates that with the aid of proposed random concentric loss, we can train an accurate semantic segmentation network in this weakly supervised application.

4.4. AMIDA 2013 dataset

We then evaluate the proposed method on the Assessment of Mitosis Detection Algorithms 2013 (AMIDA13) challenge dataset (Veta et al., 2015). This dataset consists of 12 subjects for training and 11 subjects for testing. The training set is composed of 311 HPFs and 550 annotated mitotic cells, while the test set is composed of 295 HPFs and 533 mitoses. The size of each HPF is 2000×2000 pixels, representing an area of 0.25 mm^2. Multiple pathologists annotate the HPFs, and the ground truth mitoses are the objects that have been marked by at least two experts. The label in this dataset is the centroid pixel of the mitotic cell.

We crop patches sized 500×500 pixels and perform data augmentation on them. For mitotic patches, we perform the same augmentation as that of the 2014 MITOSIS dataset. For non-mitotic patches, we rotate them with a 90-degree step. We use multiple random concentric labels for each mitotic cell. The weights of SegMitos-random model is initialized from the SegMitos-32s model that has been trained on 2012 MITOSIS dataset.

Table 6 contains the overall F-scores of our SegMitos method and some other methods. The first five methods are the participants of the AMIDA13 challenge. The IDSIA (Veta et al., 2015) method is the top ranking method, which is similar to the IDSIA (Cireşan et al., 2013) in 2012 MITOSIS challenge. CUHK and AggNet (Albarqouni et al., 2016) are the methods that submit the results after the challenge. Our SegMitos-random method surpasses all other methods remarkably, including the previous best method IDSIA (Veta et al., 2015). It is noted that the IDSIA utilizes 20 million patches extracted from training images to train three models. For improving performance, IDSIA generates eight variants for each testing image and runs the three networks on them. While our SegMitos model is trained on 300k patches and only uses the original test image without any variant.

Table 6: F-scores of different methods on AMIDA13 test set.

Method	Precision(%)	Recall(%)	F-score(%)
IDSIA (Veta et al., 2015)	61	61.2	61.1
DTU (Veta et al., 2015)	42.7	55.5	48.3
SURREY (Veta et al., 2015)	35.7	33.2	34.4
ISIK (Veta et al., 2015)	30.6	35.1	32.7
PANASONIC (Veta et al., 2015)	33.6	31.0	32.2
CUHK	**69**	31	42.7
AggNet (Albarqouni et al., 2016)	44.1	42.4	43.3
SegMitos-random	66.85	**67.73**	**67.28**

4.5. TUPAC 2016 dataset

The last dataset evaluated on is the Tumor Proliferation Assessment Challenge 2016 (TUPAC16) (Veta et al., 2018). The main goal of this challenge is to assess algorithms that predict the breast tumor proliferation from the whole-slide images. An auxiliary dataset with annotated mitotic figures is used to assess mitosis detection algorithms. The training data of this mitosis dataset consists of images from 73 breast cancer cases. The first 23 cases are the dataset that was previously released as the AMIDA13 mitosis detection challenge. The last 50 cases come from two different pathology centers, and each case has one image region. Like the other datasets, the image modification is 40X and the spatial resolution is 0.25 μm/pixel. The annotation format is the centroid pixel of mitotic cell. The testing set consists of images from 34 breast cancer cases. The ground truth of the testing data is held out by the contest organizers.

We split the training data into training set and validation set. Since the testing dataset and the cases 24-73 of the training data are from the same two pathology labs, we select validation images from the cases 24-73. Specifically, we take out one case from every seven cases as the validation set, namely, the cases 30, 37, 44, 51, 58, 65 and 72.

The size of the image region is too large so we crop patches sized 500×500

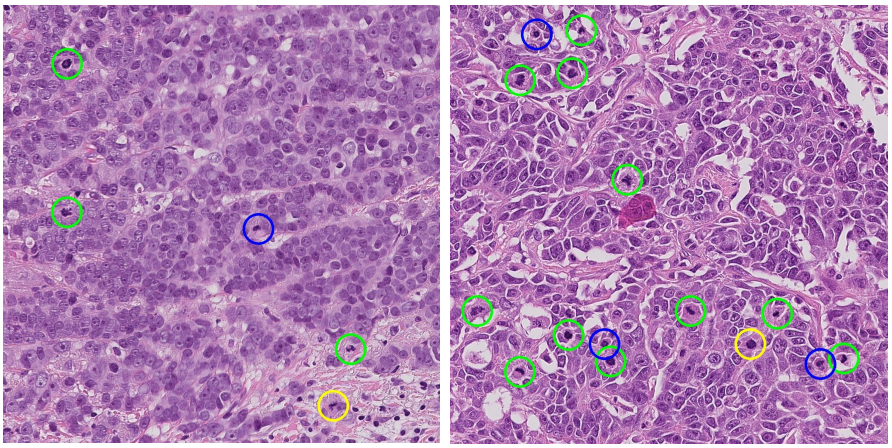


Figure 11: The detection samples of our SegMitos-random model on TUPAC16 validation set. The green, yellow and blue circles are true positives, false positives and false negatives, respectively.

pixels from the original image. We take the same method of data augmentation as that of the AMIDA13 dataset. Multiple random concentric labels are used to fit each mitotic cell.

In this dataset, our SegMitos model is fine-tuned from the PASCAL-trained FCN-32s model. No other mitosis images are used in the model training. We apply the trained SegMitos model to the validation set and achieve a 0.717 F-score. Fig. 11 shows two detection samples from the validation set.

We then add the validation set to the training set to train the SegMitos model. We utilize this model to segment the testing images of the TUPAC16 dataset and produce detection results using our screening system. We submit the results to the organizers of challenge. Table 7 contains the performance comparison of our method and other algorithms. The Lunit (Paeng et al., 2017) and the IBM (Zerhouni et al., 2017) are the top two participating approaches in the challenge. Warwick (Akram et al., 2018) achieves the best F-score among all methods after the challenge. Our method raises the F-score by 1.7% comparing with the previous best method Lunit (Paeng et al., 2017). Furthermore, our method only uses the TUPAC mitosis data while the IBM (Zerhouni et al.,

Table 7: F-scores of different methods on TUPAC16 test set. "- denotes the results which are not released.

Method	Precision(%)	Recall(%)	F-score(%)
Lunit (Paeng et al., 2017)	-	-	65.2
IBM (Zerhouni et al., 2017)	-	-	64.8
Warwick (Akram et al., 2018)	61.3	67.1	64.0
SegMitos-random	**64.0**	**70.0**	**66.9**

2017) and the Warwick (Akram et al., 2018) use some additional mitosis data.

4.6. CBIS-DDSM dataset

To demonstrate that our method can be used on other circular object segmentation, we use the concentric loss to train a breast mass segmentation model on a mammography dataset. We choose this segmentation task because many of breast masses have circle-like shapes.

We use the CBIS-DDSM dataset (Lee et al., 2017), which is an updated and standardized version of the Digital Database for Screening Mammography (DDSM)(Heath et al., 2000). The CBIS-DDSM dataset has four sub datasets: Mass-Training, Mass-Test, Calc-Training and Calc-Test. Since our target is mass segmentation, we only use the Mass-Training and Mass-test datasets. Hence, there are 1231 full mammograms in training set and 361 full mammograms in test set, respectively. Here we only regard the "Malignant" tumors as foreground object, and all "Benign" tumors are treated as background. Fig. 12 shows one full mammogram and its corresponding mask from the CBIS-DDSM training set.

Although the CBIS-DDSM dataset has provided pixel-level ROI annotations for mass, they are not generated by manually. The annotations are produced by applying a modification to the local level set framework (Chan & Vese, 1999). Though the annotations in CBIS-DDSM are much better than the DDSM-provided annotations, there still exist imprecise annotations.

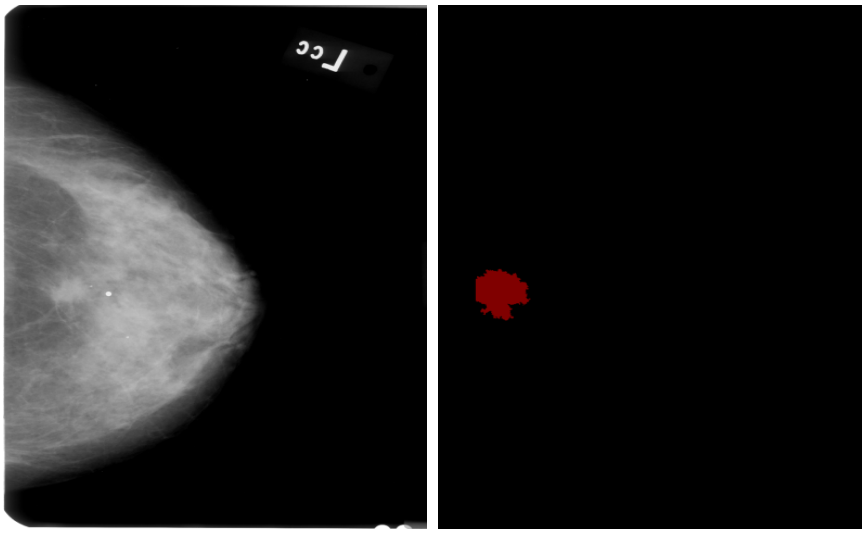


Figure 12: Mammogram and mask examples from CBIS-DDSM.

Since the original size of a full mammogram is too large to train, we first resize the images to 1536×1024 pixels and then crop the resized mammograms to obtain 6 patches. Every patch has a spatial size of 512×512 pixels. Then we perform some simple data augmentations including image flipping and rotation, yielding $19k$ image patches in total.

We produce concentric labels based on the provided pixel-level masks. As the mammographic masses are in a wide variety of shapes and sizes, we cannot use a uniform radius configuration for the concentric labels of masses. We take advantage of the height and width of each mass to generate its concentric label. The detail of producing the concentric label of a mass is shown in Fig. 13. Fig. 14 visually compares the provided pixel-level masks with the proposed concentric labels.

We use the mask labels and our concentric labels to train the FCN model, respectively. The training iterations of the two models are both $250k$. Then we run the two models on the test set of CBIS-DDSM. The dice coefficients of the two segmentation models' results are shown in Table 8. We also show some

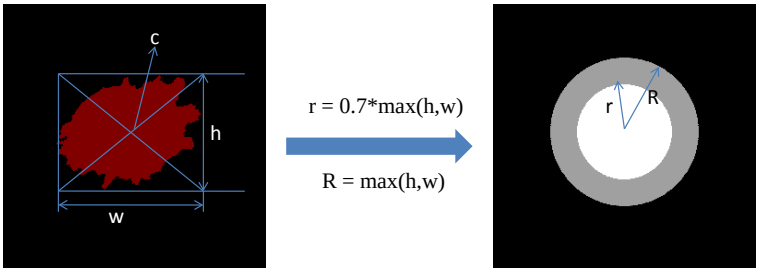
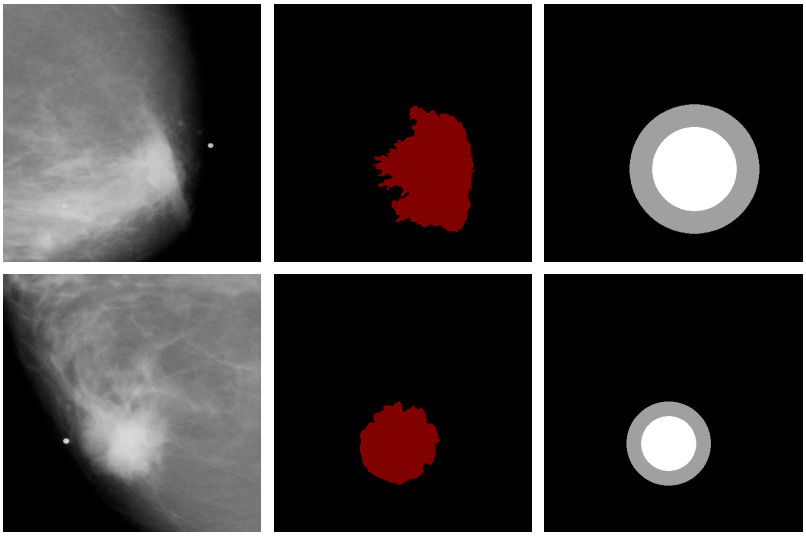


Figure 13: Illustration of generating concentric labels.



(a) mammogram (b) mask label (c) concentric label

Figure 14: Comparison of mask label and concentric label.

Table 8: Segmentation performances of FCN on CBIS-DDSM test set.

Training Labels	Dice coefficient(%)
Mask	51.41
Concentric label	50.91

segmentation results in Fig. 15.

From the Table 8, we can observe that the performance of the model trained with our concentric label is on par with that of the model trained with the original mask. This result demonstrates the effectiveness of our concentric label for this tumor segmentation task. With our concentric loss, the experts will only use a bounding box or a circle to annotate a mass, instead of manually marking every pixel along the contour of the mass.

4.7. Discussion

4.7.1. Comparison among datasets

Compared to the previous methods, the performance improvements of our methods on the three weakly supervised benchmarks (2014 MITOSIS, AMIDA13 and TUPAC16) are more obvious than that of the strongly supervised benchmark (2012 MITOSIS). That is partly because the performance on the 2012 MITOSIS dataset is already high enough due to the precise annotation. The evident improvements on the three centroid-label datasets demonstrate that our concentric loss can solve the problem of training an effective mitosis detection model with weakly supervised data. Moreover, since the image number of 2012 MITOSIS dataset is much smaller than those of the other three datasets, using the performances on the three large datasets is more convincing to compare different algorithms.

4.7.2. Comparison with humans

According to a comparison of automatic algorithms and humans for mitosis detection (Giusti et al., 2014), the IDSIA algorithm (Cireşan et al., 2013) has

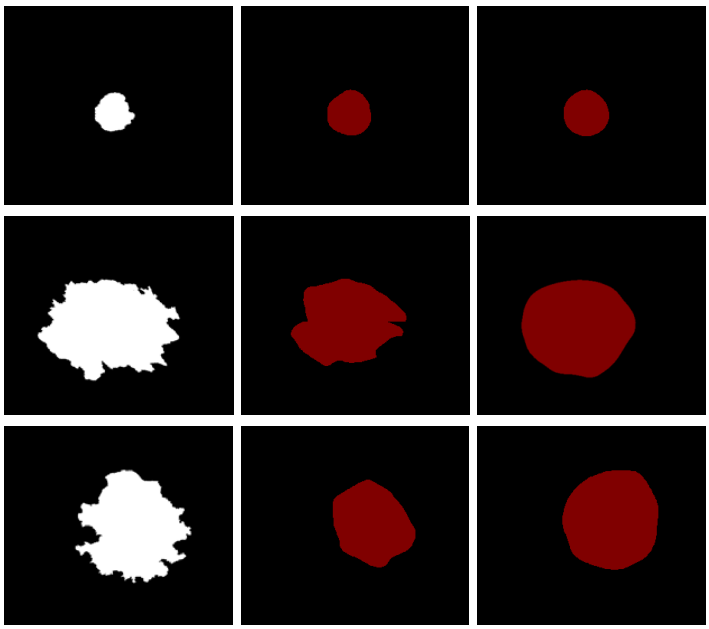


Figure 15: Segmentation results on CBIS-DDSM test set. (a) shows the mask ground truth. (b) shows the segmentation result of FCN model trained with mask labels, while (c) shows the segmentation result of FCN model trained with concentric labels.

outperformed the best test subject by a small margin in a designed classification task on 2012 MITOSIS images. In this comparison, test subjects are not previously exposed to the task and they use the same training data as the automatic algorithms. As our method has a better performance than IDSIA, this suggests that our method can also surpass the performance of humans.

4.7.3. Time analysis

In our method, segmentation of histology images occupies most of the time. For a HPF sized 2084×2084 pixels, it takes 2.9 s to produce the segmentation map. Then it takes 0.13 s to process the segmentation map and yield the final detections. Thus the total time to process a HPF on 2012 dataset is about 3 s. Compared to IDSIA (Cireşan et al., 2013) which requires 31 s to apply a network to an input image and 8 minutes to fuse results of multiple detection models for a better performance, the speed of our approach is much faster.

For the HPF in 2014 dataset, which is smaller in size (1539×1376), it takes 1.7s to process an image by our approach. In addition, we can directly deploy the SegMitos model on the entire HPF image thanks to its relatively small size. That can avoid the steps of cutting the full image and tiling the segmentation patches. Besides, deploying the SegMitos model on a full image rather than multiple patches can improve the speed. In short, processing the full HPF image directly can improve the detection speed to 0.92 s per HPF on 2014 dataset.

5. Conclusion

A deep learning based automatic and accurate mitosis detection method is proposed in this paper. We utilize a semantic segmentation model to segment breast histopathology images. Based on the probability map, we develop a filter mechanism to seek out mitotic cells from candidates. Moreover, for the weakly labeled dataset, we propose a concentric loss that can well leverage the centroid annotation to train semantic segmentation networks and obtain excellent performance. This novel loss function can extend the applicable scope

of segmentation model, especially for weakly annotated datasets. The state-of-the-art results of our SegMitos method on four mitosis datasets indicate the efficacious of deploying a segmentation model to mitosis detection task. In the future, we can apply this concentric loss to some other weakly supervised segmentation applications. Another promising future direction is to seek more powerful segmentation model on the mitosis detection task and promote the performance.

Acknowledgments

This project is sponsored by the National Natural Science Foundation of China (Grant nos. 61572207, 61733007, and 61876212), the CCF-Tencent Open Research Fund, and the Hubei Scientific and Technical Innovation Key Project. This project is further supported by the National Science Foundation Grant IIS-1302164 and IIS-1814745.

References

Akram, S. U., Qaiser, T., Graham, S., Kannala, J., Heikkilä, J., & Rajpoot, N. (2018). Leveraging unlabeled whole-slide-images for mitosis detection. *arXiv preprint arXiv:1807.11677*, .

Albarqouni, S., Baur, C., Achilles, F., Belagiannis, V., Demirci, S., & Navab, N. (2016). Aggnet: deep learning from crowds for mitosis detection in breast cancer histology images. *IEEE transactions on medical imaging*, *35*, 1313–1321.

Breiman, L. (2001). Random forests. *Machine learning*, *45*, 5–32.

Chan, T., & Vese, L. (1999). An active contour model without edges. In *International Conference on Scale-Space Theories in Computer Vision* (pp. 141–151). Springer.

Chen, H., Dou, Q., Wang, X., Qin, J., & Heng, P. A. (2016a). Mitosis detection in breast cancer histology images via deep cascaded networks. In *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence, February 12-17, 2016, Phoenix, Arizona, USA.* (pp. 1160–1166).

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

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

Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine learning*, *20*, 273–297.

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

Everingham, M., Van Gool, L., Williams, C. K., Winn, J., & Zisserman, A. (2010). The pascal visual object classes (voc) challenge. *International journal of computer vision*, *88*, 303–338.

Freund, Y., & Schapire, R. E. (1995). A desicion-theoretic generalization of on-line learning and an application to boosting. In *European conference on computational learning theory* (pp. 23–37). Springer.

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).

Giusti, A., Caccia, C., Cireşari, D. C., Schmidhuber, J., & Gambardella, L. M. (2014). A comparison of algorithms and humans for mitosis detection. In *Biomedical Imaging (ISBI), 2014 IEEE 11th International Symposium on* (pp. 1360–1363). IEEE.

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).

Heath, M., Bowyer, K., Kopans, D., Moore, R., & Kegelmeyer, W. P. (2000). The digital database for screening mammography. In *Proceedings of the 5th international workshop on digital mammography* (pp. 212–218). Medical Physics Publishing.

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

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

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

LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, *86*, 2278–2324.

Lee, R. S., Gimenez, F., Hoogi, A., Miyake, K. K., Gorovoy, M., & Rubin, D. L. (2017). A curated mammography data set for use in computer-aided detection and diagnosis research. *Scientific data*, *4*, 170177.

Li, C., Wang, X., & Liu, W. (2017). Neural features for pedestrian detection. *Neurocomputing*, *238*, 420–432.

Li, C., Wang, X., Liu, W., & Latecki, L. J. (2018). Deepmitosis: Mitosis detection via deep detection, verification and segmentation networks. *Medical image analysis*, *45*, 121–133.

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

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

MITOS-ATYPIA-14 (2014). Mitos-atypia-14-dataset. `https://mitos-atypia-14.grand-challenge.org/dataset/`. [Online; accessed 17.12.14].

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

Paeng, K., Hwang, S., Park, S., & Kim, M. (2017). A unified framework for tumor proliferation score prediction in breast histopathology. In *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support* (pp. 231–239). Springer.

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

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

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

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

Shen, W., Wang, B., Jiang, Y., Wang, Y., & Yuille, A. (2017a). Multi-stage multi-recursive-input fully convolutional networks for neuronal boundary detection. In *2017 IEEE International Conference on Computer Vision (ICCV)* (pp. 2410–2419). IEEE.

Shen, W., Zhao, K., Jiang, Y., Wang, Y., Bai, X., & Yuille, A. L. (2017b). Deepskeleton: Learning multi-task scale-associated deep side outputs for object skeleton extraction in natural images. *IEEE Transactions on Image Processing*, *26*, 5298–5311.

Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, .

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

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., & Rabinovich, A. (2015). Going deeper with convolutions. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 1–9).

Tajbakhsh, N., Shin, J. Y., Gurudu, S. R., Hurst, R. T., Kendall, C. B., Gotway, M. B., & Liang, J. (2016). Convolutional neural networks for medical image analysis: full training or fine tuning? *IEEE transactions on medical imaging*, *35*, 1299–1312.

Tang, P., Wang, X., Bai, S., Shen, W., Bai, X., Liu, W., & Yuille, A. L. (2018). Pcl: Proposal cluster learning for weakly supervised object detection. *IEEE transactions on pattern analysis and machine intelligence*, .

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.

Tang, P., Wang, X., Shi, B., Bai, X., Liu, W., & Tu, Z. (2016). Deep fishernet for object classification. *arXiv preprint arXiv:1608.00182*, .

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

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

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

Veta, M., Heng, Y. J., Stathonikos, N., Ehteshami Bejnordi, B., Beca, F., Wollmann, T., Rohr, K., Shah, M. A., Wang, D., Rousson, M., Hedlund, M., Tellez, D., Ciompi, F., Zerhouni, E., Lanyi, D., Viana, M., Kovalev, V., Liauchuk, V., Ahmady Phoulady, H., Qaiser, T., Graham, S., Rajpoot, N., Sjöblom, E., Molin, J., Paeng, K., Hwang, S., Park, S., Jia, Z., I-Chao Chang, E., Xu, Y., Beck, A. H., van Diest, P. J., & Pluim, J. P. W. (2018). Predicting breast tumor proliferation from whole-slide images: the TUPAC16 challenge. *ArXiv e-prints*, . `arXiv:1807.08284`.

Veta, M., Van Diest, P. J., Willems, S. M., Wang, H., Madabhushi, A., Cruz-Roa, A., Gonzalez, F., Larsen, A. B., Vestergaard, J. S., Dahl, A. B. et al.

(2015). Assessment of algorithms for mitosis detection in breast cancer histopathology images. *Medical image analysis*, *20*, 237–248.

Wang, H., Cruz-Roa, A., Basavanhally, A., Gilmore, H., Shih, N., Feldman, M., Tomaszewski, J., Gonzalez, F., & Madabhushi, A. (2014). Cascaded ensemble of convolutional neural networks and handcrafted features for mitosis detection. In *SPIE Medical Imaging* (pp. 90410B–90410B). International Society for Optics and Photonics.

Wang, X., Yan, Y., Tang, P., Bai, X., & Liu, W. (2018). Revisiting multiple instance neural networks. *Pattern Recognition*, *74*, 15–24.

Wang, X., Yang, W., Weinreb, J. C., Han, J., Li, Q., Kong, X., Yan, Y., Ke, Z., Luo, B., Liu, T., & Wang, L. (2017). Searching for prostate cancer by fully automated magnetic resonance imaging classification: deep learning versus non-deep learning. In *Scientific Reports*.

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

Yin, S., Peng, Q., Li, H., Zhang, Z., You, X., Furth, S. L., Tasian, G. E., & Fan, Y. (2018). Subsequent boundary distance regression and pixelwise classification networks for automatic kidney segmentation in ultrasound images. *arXiv preprint arXiv:1811.04815*, .

Zerhouni, E., Lányi, D., Viana, M., & Gabrani, M. (2017). Wide residual networks for mitosis detection. In *Biomedical Imaging (ISBI 2017), 2017 IEEE 14th International Symposium on* (pp. 924–928). IEEE.

Zhou, Y., Xie, L., Shen, W., Wang, Y., Fishman, E. K., & Yuille, A. L. (2017). A fixed-point model for pancreas segmentation in abdominal ct scans. In *International Conference on Medical Image Computing and Computer-Assisted Intervention* (pp. 693–701). Springer.