Few-shot Breast Cancer Metastases Classification via Unsupervised Cell Ranking

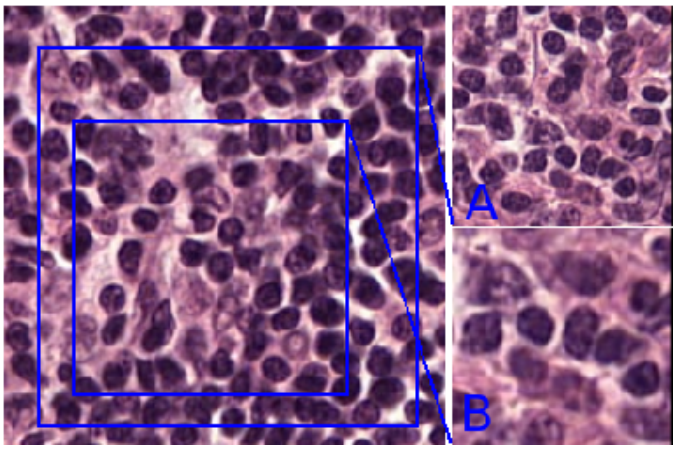
Jiaojiao Chen, Jianbo Jiao, Shengfeng He, Guoqiang Han, and Jing Qin

***Abstract*—Tumor metastases detection is of great importance for the treatment of breast cancer patients. Various CNN (Convolutional Neural Network) based methods get excellent performance in object detection/segmentation. However, the detection of metastases in hematoxylin and eosin (H&E) stained whole-slide images (WSI) is still challenging mainly due to two aspects: (1) The resolution of the image is too large. (2) lacking labeled training data. Whole-slide images generally stored in a multi-resolution structure with multiple downsampled tiles. It is difficult to feed the whole image into memory without compression. Moreover, labeling images for the pathologists are time-consuming and expensive. In this paper, we study the problem of detecting breast cancer metastases in the pathological image on patch level. To address the abovementioned challenges, we propose a few-shot learning method to classify whether an image patch contains tumor cells. Specifically, we propose a patch-level unsupervised cell ranking approach, which only relies on images with limited labels. The main idea of the proposed method is that when cropping a patch A from the WSI and further cropping a sub-patch B from A, the cell number of A is always larger than that of B. Based on this observation, we make use of the unlabeled images to learn the ranking information of cell counting to extract the abstract features. Experimental results show that our method is effective to improve the patch-level classification accuracy, compared to the traditional supervised method. The source code is publicly available at https://github.com/fewshot-camelyon.**

***Index Terms*—Few-shot Learning, Metastases Classification, Unsupervised Learning**

I. Introduction

Cancer is a disease that cells divide without stopping. In general, the earlier cancer diagnosed, the more powerful effect the treatment could be. A complete process of breast cancer diagnose on pathological images contains cancer classification/detection [1], [2], grading(cell analysis) [3], [4], and staging (the extent of the disease) [5], [6], [7], according to the TNM (Tumor, Node, Metastasis) staging system. In this paper, we study the breast cancer whole-slide pathological image of lymph node sections that is adjacent to the breast, to detect the breast cancer metastases.



$C(A) > C(B)$

Fig. 1: The main idea of our work. The right images are cropped patches from the left one, in the corresponding square regions. We first crop a patch A from the original slide image. Then we further crop a sub-patch B from A. It can be observed that the number of cells in A is always larger than that in B. By setting patch A as the positive sample and B the negative sample, we learn a binary classifier to extract the ranking features, in a unsupervised manner. Consequently, a binary classifier is fine-tuned based on the above features to detect the metastases.

Before machine learning methods appear, pathologists diagnose cancer metastases by scanning the pathological images of the patients under the microscopes. However, it is time-consuming and even the skilled pathologists are still error-prone. Moreover, skilled pathologists are scarce especially for those remote areas. With the development of machine learning, various CNN-based methods are proposed to solve the problems of image recognition [8], [9], [10], image semantic segmentation [11], [12], [13], *etc.*, which show promising performance. In medical imaging domain, there are also many attempts of applying machine learning methods on pathological images. Moreover, with the great progress of machine learning, using auto-classification [14], [15], [16] and detection methods [17], [18] to analyse breast cancer has become the mainstream. All these methods are time-saving and labor-saving. However, almost all these methods are data-driven which relies on a large number of labeled training data. As discussed above, it is difficult and time-

This project is supported by the National Natural Science Foundation of China (No. 61472145, No. 61972162, and No. 61702194), a grant from the Hong Kong Research Grants Council (Project No. PolyU 152035/17E), the Special Fund of Science and Technology Research and Development on Application From Guangdong Province (SF-STRDA-GD) (No. 2016B010127003), the Guangzhou Key Industrial Technology Research fund (No. 201802010036), the Guangdong Natural Science Foundation (No. 2017A030312008), and the CCF-Tencent Open Research fund (CCF-Tencent RAGR20190112). Corresponding author: Shengfeng He.

Jiaojiao Chen, Shengfeng He and Guoqiang Han are with the School of Computer Science and Engineering, South China University of Technology, Guangzhou, China. E-mail: cjiaojiao9494@gmail.com, {hesfe, csgqhan}@scut.edu.cn.

Jianbo Jiao is with the Department of Engineering Science, the University of Oxford. Email: jianbo@robots.ox.ac.uk.

Jing Qin is with the Department of Nursing, Hong Kong Polytechnic University. E-mail: harry.qin@polyu.edu.hk.

consuming for pathologists to scan and label the medical images, especially for the gigapixel pathology images. In order to deal with the problem of lacking training data, some deep learning methods [19], [20] train on other larger datasets and finetune with the target one. Other methods [21], [15] use data augmentation strategies to expand the dataset. Chang *et al.*[19] propose a transfer learning method to detect breast cancer, which trained on non-medical images first and transferred to the histopathology images. Samala *et al.*[20] develop a deep CNN using a multi-stage transfer learning approach that utilized data from similar auxiliary domains. Bayramoglu *et al.*[21] use multi-scale images as inputs of CNN, and Motlagh *et al.*[15] use random resizing, rotating, cropping and flipping methods to augment the breast cancer sub-type classes. In a word, abundant dataset or other data expansion methods are used in the works described above.

In this paper, we propose a few-shot classification method that utilizes cell counting as an auxiliary task. The main idea is that even though the exact counting C(A) of a cell image A is unknown, the counting C(B) of a sub-image B (cropped from A) is less or equal to C(A) (see Figure 1). In our work, we do not have abundant labeled data for training, but we can get rich unlabeled samples from the hospital and other organizations. So in our task, we crop a pair of patches from the original input and resize them to the same size. We can know that the cell number of the small patch is less or equal to the original (Figure 1). The auxiliary ranking task can learn more abstract features and boost the performance of the supervised classification method with insufficient data.

In summary, the main contributions of this paper are three-fold.

- We address the breast cancer metastases classification problem in an unsupervised manner.
- We propose a few-shot learning method that takes cell ranking as an auxiliary task.
- The proposed approach is shown to perform better than the traditional supervised methods trained with the same amount of labeled data.

The rest of the paper is organized as follows. In the next section we briefly describes the recent work of cancer metastases classification and detection. In section III, we describe the proposed method in detail. In section IV we show the performance of our method with extensive experiments. Finally in section V, we conclude our work.

II. Related Work

Traditional machine learning method. Breast cancer is the most common cancer among women. Before machine learning appears, pathologists judge breast cancer by scanning the pathologist images through the microscopes [22]. Later on, machine learning based methods [23], [24] appear to perform computer assisted diagnosis. In [23], the authors present naive Bayes classifier and k-neareast neighbor classifier for breast cancer classification. Random Forests classifier with feature combination of Gabor and gray-level co-occurrence matrix methods is used in [14] to perform texture analysis. Support vector machine is used in [24] for breast cancer classification

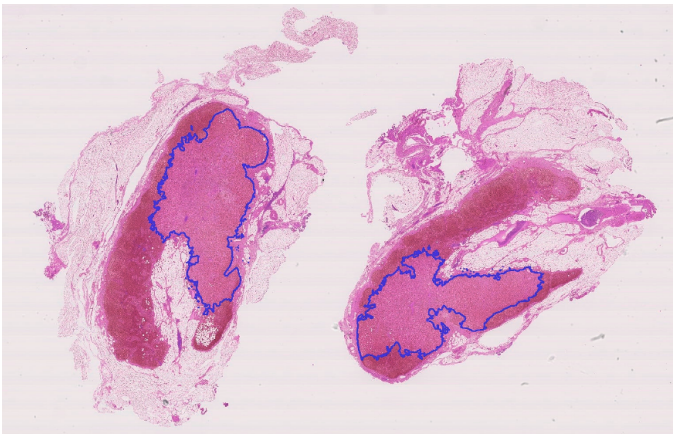


Fig. 2: The seventh downsampled tile of the lymph node pathology tissue. The cells marked with the blue line are tumor metastases.

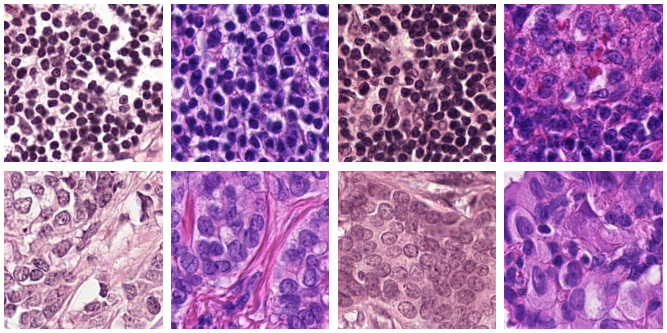


Fig. 3: Some normal (top) and tumor (bottom) patches cropped from the whole-slide pathological tissue in the biggest tile.

with gene expression data. However, these methods use hand-crafted features which cannot fully represent the intrinsic properties of breast cancer histology images.

Deep learning method. Recently, convolution neural networks have achieved great success in object classification and detection. There are also some attempts [21] [18], [20], [25], [26] proposeed to deal with breast cancer classification and detection. Bayramoglu *et al.*[21] propose to use CNNs to classify breast cancer histopathology images independent of their magnification. Spanhol *et al.*[16] propose a method based on the extraction of image patches for training the CNN to classify breast cancer. This method allowed using the high resolution histopathological images as input, and the performance is better when compared to other machine learning models trained with hand-crafted image descriptors. Deep max-pooling CNNs is proposed in [4] to detect mitosis in breast cancer histology images. Aresta *et al.* [27] propose to classify normal, benign, in situ carcinoma and invasive carcinoma of breast cancer on the whole-slide level. A method with anchor layers scanning the whole-slide image to predict micro- and macro-metastases is proposed in [28]. van Diest *et al.* [29] summarize the performance of deep learning algorithms for detecting metastases in H&E images of lymph of women with breast cancer. Wang *et al.*[30] achieve the

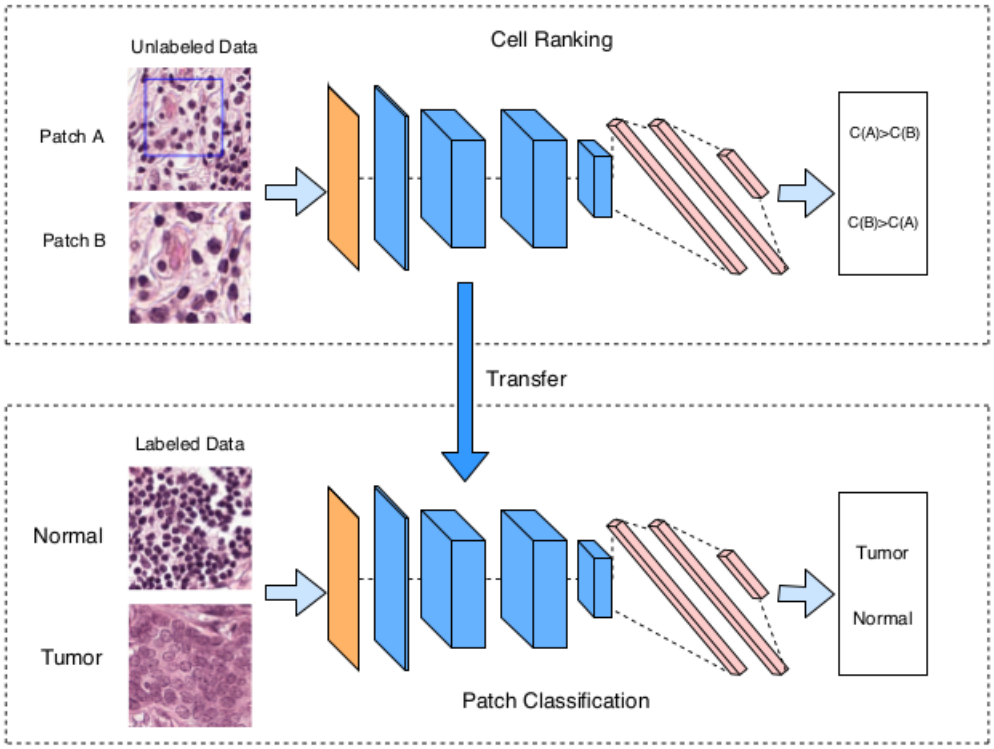


Fig. 4: The framework of our few-shot classification method. The main procedure contains ranking and classification. The ranking step learns from cell ranking with the pairs of patches cropped from the unlabeled WSI. The unsupervised ranking knowledge is then transferred to the classification network, followed by a finetuning step using the labeled training data cropped from the labeled WSI in a supervised manner.

best performance on CAMELYON16[1] (Cancer Metastasis in Lymph Nodes Challenge 2016) for breast cancer detection. The result in [31] shows that categories of false positives, *e.g.* , nerves or contamination, could be further optimized. These methods achieved promising results in breast cancer classification and detection. However, they rely heavily on large amount of annotated training data.

Semi-supervised learning method. To deal with the problem of lacking training data, some few-shot learning methods [32], [33] are proposed. In [32], SVM classifier is employed to mask unlabeled data, followed by a decision tree that is learned with masked unlabeled data and real-labeled data for diagnosis of ultrasound breast tumor. Oliver *et al.*[33] propose a one-shot learning method to segment the breast, pectoral muscle, and background in digitized mammograms with position, intensity and texture information. On the other hand, semi-supervised method is used in [34], [35], [36]. Sun *et al.*[35] make full use of the unlabeled data and proposed a graph based semi-supervised learning scheme using deep conventional neural network for breast cancer diagnosis. Li *et al.*[36] present a semi-supervised locality discriminant projections with kernels for breast cancer classification. However, the above mentioned methods either fine-tuned with large dataset that unrelated to the origin, or reply on data expansion method. In our work, we leverage limited labeled data and make full use of the unlabeled original data.

[1]https://camelyon16.grand-challenge.org/

III. Few-shot Cancer Metastases Classification

A. Overview

Our work aims to detect the breast cancer metastases with the gigapixel pathology images of lymph node sections adjacent to the breast. As discussed in the introduction part, the whole-slide images are stored in a multi-resolution structure with multiple downsampled tiles which are hard to feed into the memory. As a result, our work is carried out on the patch-level and the patch is cropped at the largest layer. The whole work contains two steps: 1) training patch-level classifier. 2) Detection (mask generation). Based on the patch-level classification of whether a patch contains tumor, we can generate the tumor metastases area of the whole-slide image. Figure 2 shows an example of the seventh down-sampled layer of the pathology image. Some example patches are shown in Figure 3. In our patch-level classification, we use the tumor/normal patches cropped from the labeled WSI to train a binary classifier. As mentioned above, acquiring labeled data in the medical domain is costly. Although supervised learning is available using the patches cropped from the labeled whole-slide tissue, it is difficult to get satisfied result because of limited data. While the unlabeled data is relatively easier to obtain. So in order to get more representative features, we take an auxiliary unsupervised cell ranking method which

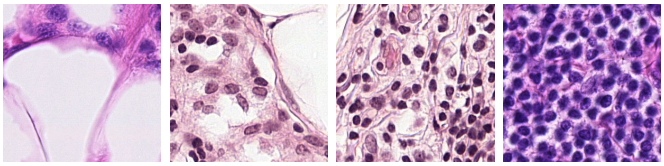


Fig. 5: Example cell images with different dense levels.

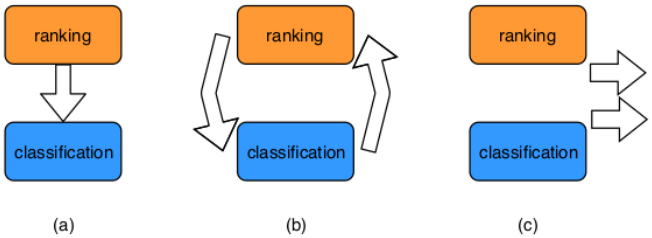


Fig. 6: Three methods of combining the unsupervised ranking task and the patch-level classification task. (a) learns ranking task and classification task in sequential order. (b) learns two tasks iteratively. (c) learns two tasks simultaneously.

proposed to use the unlabeled dataset to improve the patch-level classification accuracy. So training patch-level classifier contains two steps: 1) Supervised patch classification with tumor/normal patches cropped from the WSI. 2) Unsupervised cell ranking with pairs of patches cropped from the unlabeled WSI. Figure 4 shows the pipeline of the proposed method.

B. Supervised Patch Classification

In the supervised patch-level classification, the inputs are patches cropped from the gigapixel pathology whole-slide images with a sliding window, and outputs are the probability of normal or tumor class the patches belong to. It is a binary classification problem and cross entropy loss is used as the training objective. In our work, we use the Inception [37] as our classification network. In addition, we also utilize the VGG [38] and ResNet [39] to verify the effectiveness of our method. Specifically, we choose resnet50 and vgg19 for our verification work. Both the ranking task and the classification task are binary classification problems. And supervised patch classification is shown in the bottom of Figure 4. For the supervised classification phase, the loss function is defined as:

$$Loss_c = -(1/N)\sum_{i=1}^{N}\sum_{j=1}^{M} y_{ij}\log(p_{ij}), \tag{1}$$

where N denotes the number of training samples, M denotes the number of classes. p_{ij} is the probability of image i belongs to class j, while y_{ij} is the ground truth probability.

C. Unsupervised Cell Ranking

Learning an automatic diagnose system in a supervised manner with limited data is a challenging problem. Our main idea is to use unlabeled data for cell counting by learning to rank. It is natural for an image patch within a larger patch should have fewer or equal number of elements compared to the larger one. In our scenario, when cropping a patch from a larger pathology patch, the cell counting of the larger patch is bigger than the small patch (Figure 1). Figure 5 shows some cell images in which each patch has a different number of cells. In the phase of unsupervised cell ranking, we assign a positive label one to the larger patch while a negative label to the smaller one. These patches are then used to train a binary classifier. The framework of the unsupervised cell ranking is shown in the top of Figure 4. To present the model from learning only the magnification information, we randomly crop the larger patch within the width range between 400 and 800, and the smaller patch with a random ratio between 0.4 to 0.9 of the bigger one. Then all the patches are resized to the same size. By this way, each pair of patches are independent, although the small patch selected from one pair A may have more cells than another big patch selected from pair B. In the ranking phase, the loss is defined as:

$$Loss_r = -(1/2K)\sum_{i=1}^{K}\sum_{j=1}^{L} y_{ij}\log(p_{ij}), \tag{2}$$

where K denotes the number of training pairs, L denotes the positive and negative class. p_{ij} predicted the probability of patch i belongs to class j, and y_{ij} means the ground truth that patch i belongs to class j.

There are three possible ways to combine the ranking task and supervised classification task (Figure 6): 1) Do cell ranking first and classification next. 2) Do two task iteratively or 3) two task learn simultaneously. Learning these two tasks simultaneously with weight sharing requires large amount of data in both tasks, which is not applicable for few shot learning. On the other hand, iteratively training two tasks can easily lead to overfitting. As a result, in our work, we model our method as a sequential one. We use the unsupervised cell ranking task as pre-training stage to extract common features, then finetune the low-level features and the final classifier.

D. Detection (Mask Generation)

Different from the natural image detection or localization methods that input the whole image into network, we feed small patches into our classification network. These small patches are extracted in the sliding window fashion on the pathology image. Then we combine the predicted results of those small patches into a two-dimensional matrix according to the order. The matrix is further transformed into a heatmap. The final detection result is achieved by comparing the heatmap with the ground truth binary map labeled by the pathologist. Some post-processing techniques are also applied in the heatmap, *e.g.* removing those probabilities less than 0.5 in the two-dimensional matrix and merge two tumor areas that are too closed in the heatmap. The framework of mask generation is shown in Figure 7.

Next, we describe how to generate the heatmap and binary diagram in detail. The method is illustrated in Figure 8. The pathological images are stored with multiple down-sampled layers of scale 2 spatially. Our detection result is generated in

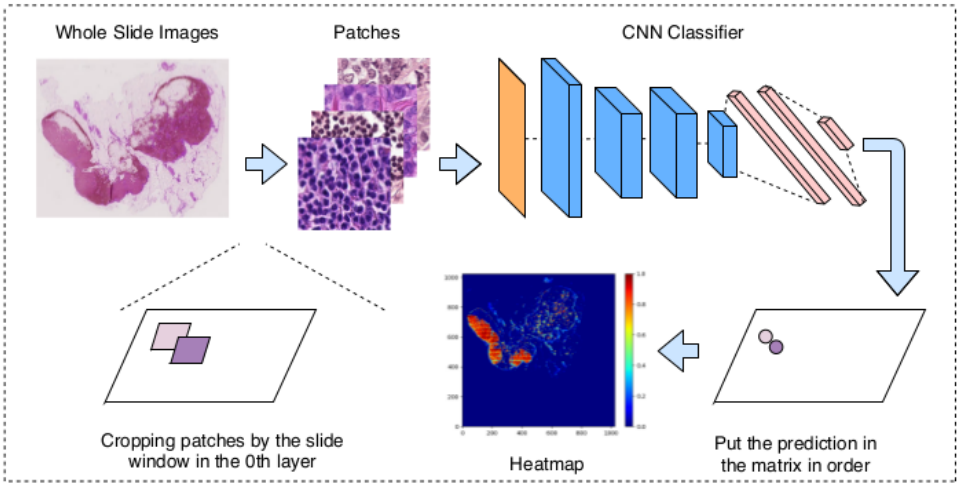


Fig. 7: The framework of metastases detection (mask generation). Patches are first extracted from the whole-slide tissue image with a sliding window. Then the patches are fed into a pre-trained CNN model. A two dimensional matrix is generated from the prediction of the CNN model and finally the detection mask is achieved from the matrix. The CNN model is the patch-level classifier and the step length of the sliding window with overlapping is 128.

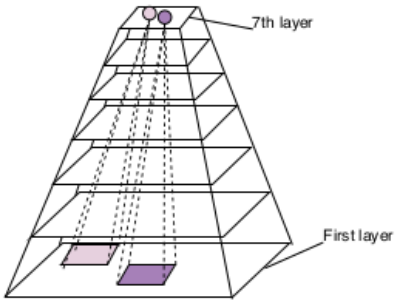


Fig. 8: Patches cropped by the sliding window in detail and the heatmap generated from the two-dimensional matrix. Each point in the 7th layer corresponds to a patch in the first layer.

the seventh layer, and we use a sliding window with a step length of 128 to crop patch (with overlapping). Each point in the 7th layer corresponds to a patch of 128*128 in the first layer. Different networks with different input sizes but all the input we used is larger than 128. As shown in Figure 8, when cropping a patch with sliding window and feed it into the classifier, the classifier outputs the predicted probability of tumor and fill it in the corresponding position of the matrix with the same size of the 7th layer.

IV. Experiments

A. Implementation Details

Our experiments are performed on the CAMELYON16 dataset. It is collected from two independent breast cancer datasets which contain 270 training WSI and 130 test WSI. We use the patch-level classification accuracy as our criterion. We randomly selected two tumor samples (because it contains not only tumor cells but also normal cells) and the corresponding masks from the training set as our training data. With the samples selected, we crop approximately 30k training samples (the normal samples and tumor sample is at a one-to-one scale) and 20k validation samples. We then use the other 266 images selected from the CAMELYON dataset as the unlabeled dataset and crop approximately 60k pairs of patches for learning the ranking task. In order to evaluate the effectiveness of our work, we test the proposed method using the testing set of CAMELYON, which contains 130 images (we crop approximately 3.8 million patches). In each ranking epoch, 2k batches (batch size of 16) are randomly selected for training and validation. We trained our networks with stochastic gradient descent in Pytorch and the loss function is cross entropy loss. Unless otherwise stated, the weight decay is 0.05. To combine the classification and the ranking modules, we first pre-train on the ranking data to extract common features, and then finetune the model on the classification data. In the finetuning phase, the learning rate of the fully connected layers is ten times of the convolutional layers. The training data distribution is described in Table I and the training parameters are described in Table II.

TABLE I: The number of whole-slide images and patches cropped from the WSI about the supervised learning with the labeled samples and the data of cells counting for ranking. The number of patches for ranking is in pairs.

Num. \ Phase	Train	Val	Test	Rank
num. of WSI	2	2	130	266
patches	30925	19799	3.8m	60k

B. Ablation Studies and Quantitative Evaluations

In this section, we perform an ablation study on the CAMELYON16 dataset. We first compare the proposed few-shot classification with state-of-the-art fully-supervised method. The results are shown in Table III. The work by Wang *et*

TABLE II: Parameters details. This table shows the learning rate and batch size used in the ranking and classification phase.

Method \ Model	Resnet	VGG	Inception
learning rate(lr)	1e-05	1e-05	1e-04
ranking lr	5e-04	1e-04	1e-03
batch size(bs)	32	32	32
ranking bs	12	10	16

al.[30] is the winner of the camelyon16 challenge. They use all the 270 labeled whole-slide images to perform a fully-supervised learning and trained with multiple GPUs. We see that the proposed method is trained with only 2 labeled images, while still achieves reasonable results compared to the state-of-the-art fully-supervised method.

We also compare the differences between the proposed approach that uses the auxiliary task with training from scratch, and we use training from scratch in a supervised manner as the baseline. The results are shown in Table IV. We can see that for each model, there is an significant improvement for patch-level classification accuracy. It validates the effectiveness of the proposed ranking task in our work.

After comparing the performance of training from scratch and pre-trained with the ranking information, we also analyze the influence of ranking with different training dataset size. For this evaluation, we get two large and small training datasets of approximately 100k and 5k patches for training (compared to the original 30k patches). The validation data size and test data size are 20k and 3.8m, respectively. The result is shown in Table V. From the table we can see that even for a much small training data, ranking task provides an obvious improvement of classification accuracy. It reveals the same conclusion of the ablation studies. In addition, we can see that the improvement of the 30k data size is larger than that of the 5k size. This is because the small dataset has much less training samples that are insufficient for learning representative features. On the other hand, using a large data with 100k patches can obtain a much better performance, and the ranking strategy still boosts the classification accuracy for such a large number of training samples. All other settings are similar, except the data size for training between three experiments.

Finally, we compare the performance of our proposed method with the leaderboard[2] of the CAMELYON16 challenge. The CAMELYON16 challenge contains two tasks: *Identification of individual metastases in WSI* and *Classification of slides containing metastases and normal sildes*.

Identification of individual metastases in WSI: In this task, free-response receiver operating characteristic (FROC) curve is used. The FROC curve is defined as the plot of true-positive fraction versus the average number of false-positive per WSI. The scoring matrix of CAMELYON16 is defined as the average sensitivity at 6 predefined false positive rates: 1/4, 1/2, 1, 2, 4 and 8 false positives per WSI. At this stage, all the probability heatmaps of test set are generated using the trained patch-level CNN, and binarized with a threshold value of 0.9. Finally, all the predicted regions are post-processed with morphological operators to connect neighboring regions. The centroids of the connected regions are selected as the candidate lesion location and the probability of the centroids are used as the lesion scores. The result is shown in Figure 9. We can see that our method ranks at 11th among 32 algorithms with just 2 pixel-level labeled WSI. Moreover, we can see that the scores of each algorithm varied significantly despite the use of advanced CNN architectures. As discussed in paper[29], the classification performance is largely influence by different auxiliary strategies (*e.g.* , standardization technique, network ensemble and hard-negative mining). In here we do not further delve into these tricks, but to only show the effectiveness of our ranking model. However, the proposed few-shot model achieves reasonable results, and even outperforms some deep learning based fully supervised model (*e.g.* , DeepCare).

Classification of slides containing metastases and normal sildes: For this task, a single probability of tumor for the entire WSI is predicted. At this stage, *HMS & MIT (I & II)* extract a 28-length feature vectors from the heatmaps to train a random forest classifier, and use this classifier to assign the slide level score. *HMS & MGH(I, II & III)* are also extracted higher level features from the heatmaps to train a random forest classifier, and we use the classifier to produce the probability score for each slide. Some teams use the maximum lesion score in the first task as the slide score like *CULab(I, II & III) and ExB*. In our work, we also extract 30 geometrical and morphological features from the heatmaps of all training data. They include the total number of tumor regions per WSI, percentage of tumor region over the whole tissue region, the area of largest tumor region, the longest axis in the largest tumor region, average prediction across tumor region, max, mean, variance, skewness, kurtosis of area, perimeter, compactness, rectangular and solidity. We use these features to build a random forest classifier to predict the WSI score of the test set. The result is shown in Figure 10. On the test phase, our slide-level classification method achieves an AUC of 0.883.

C. Qualitative Evaluation

In this part, we qualitatively compare our method with the baseline and the labeled ground truth segmentation. As shown in Figure 11, the third column is the results of our cancer segmentation, while the fourth column is the baseline that is trained from scratch on the patch-level. In the segmentation results, blue region indicates normal and background, while red means tumor. The color between red and blue means the different probability of tumor. From the result, we can see that our method can well detect the metastases, and the proposed method performs much better than the baseline. For example, it is obviously that our method has less noise then the baseline in the ambiguous regions. It also indicates that the proposed learning to rank is effective for learning representative features.

D. Limitation and Discussion

In our work, we proposed a method that utilizes the easily accessible unlabeled data, to promote the breast cancer metastases segmentation with limited labeled data. The proposed

[2]https://camelyon16.grand-challenge.org/Results/

TABLE III: Patch-level classification accuracy of fully-supervised method used in the challenge winner [30] and our method.

Method	few-shot			fully-supervised	
	Resnet50	VGG19	InceptionV3	VGG16	InceptionV1
[30]	-	-	-	0.9790	0.9840
Ours	0.8601	0.8720	0.8961	-	-

TABLE IV: Evaluation on the patch-level. The numbers are the final patch-level classification accuracy of the proposed method with different models on the testset. Base CNN means train from scratch in the training set cropped from the two whole-slide images. Finetune means pretraining on the ranking dataset and finetuning it with the training dataset. Improvement means the efficiency of using ranking task compared to the base CNN.

Method \ Model	Resnet	VGG	Inception
Base CNN	0.7965	0.8040	0.8680
Finetune	0.8601	0.8720	0.8961
Improvement	0.0636	0.0680	0.0281

TABLE V: Performance improvement of patch-level classification accuracy of the proposed method with different size of training data. 5k, 30k, and 100k indicate the training samples used for training. The first line of each size means train from scratch, while the second line means pre-trained with learning to rank.

data size \ Model	Resnet	VGG	Inception
100k samples	0.9012	0.9115	0.9356
	0.9305	0.9401	0.9522
Improvement	0.0293	0.0286	0.0166
30k samples	0.7965	0.8040	0.8680
	0.8601	0.8720	0.8961
Improvement	0.0636	0.0680	0.0281
5k samples	0.8034	0.8185	0.8571
	0.8763	0.8688	0.8752
Improvement	0.0639	0.0503	0.0181

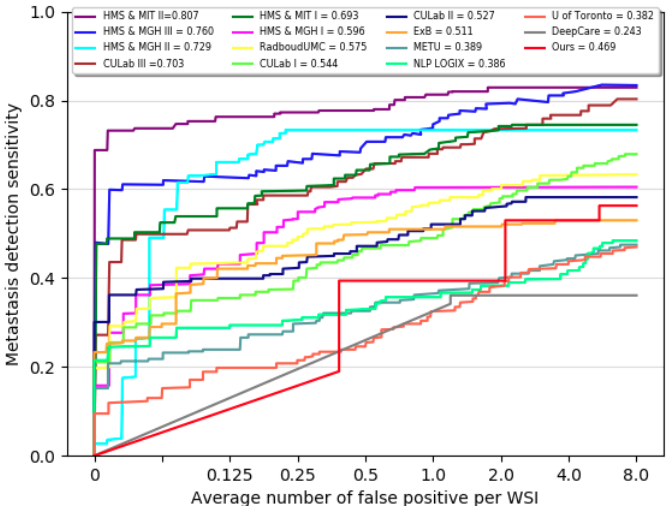


Fig. 9: FROC curves of the Top algorithms for the Metastases Identification Task From the CAMELYON16 Competition.

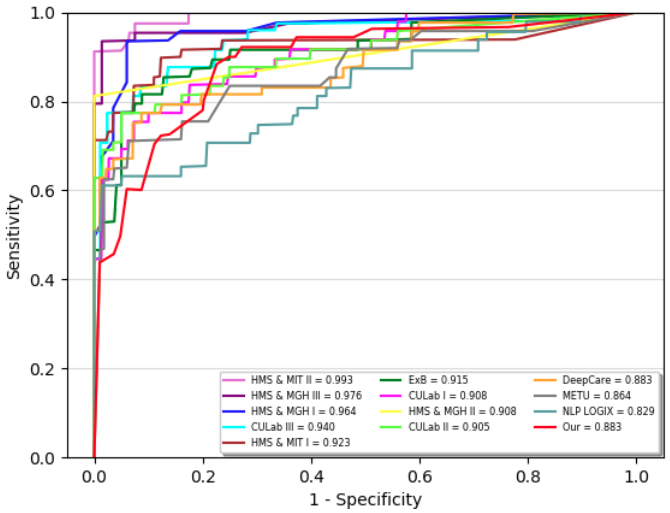


Fig. 10: ROC curves of the Top Algorithms for the Metastases Classification Task From the CAMELYON16 Competition.

method achieved a better result compared to traditional supervised learning methods and got closed to the fully-supervised method training with enough data. One limitation of our work is that some noise exists in the result, mainly caused by the lacking of training data and inadequate annotations. As shown in Figure 12, although our method has less noise then the baseline, there are also much uncertain like the yellow area. The reason is the false prediction of the patch-level classifier, which may be caused by the new type of tumor cells that have not appeared in the training phase. Another limitation of our work is that for small metastases area, especially for isolated tumor cells, it is difficult to detect (*e.g.* , Figure 13). In Figure 13, the area contoured by the red line have not been detected. We may focus on these two main limitations in our future work, by proposing a better solution to incorporate unlabeled data. On the other hand, our learning to rank idea is general and can be applied to other types medical images. It will leaves as one of our future works.

V. Conclusion

In this paper, we proposed a few-shot learning method to detect the cancer metastases on patch-level. Our method takes full advantage of the unlabeled data to learn ranking

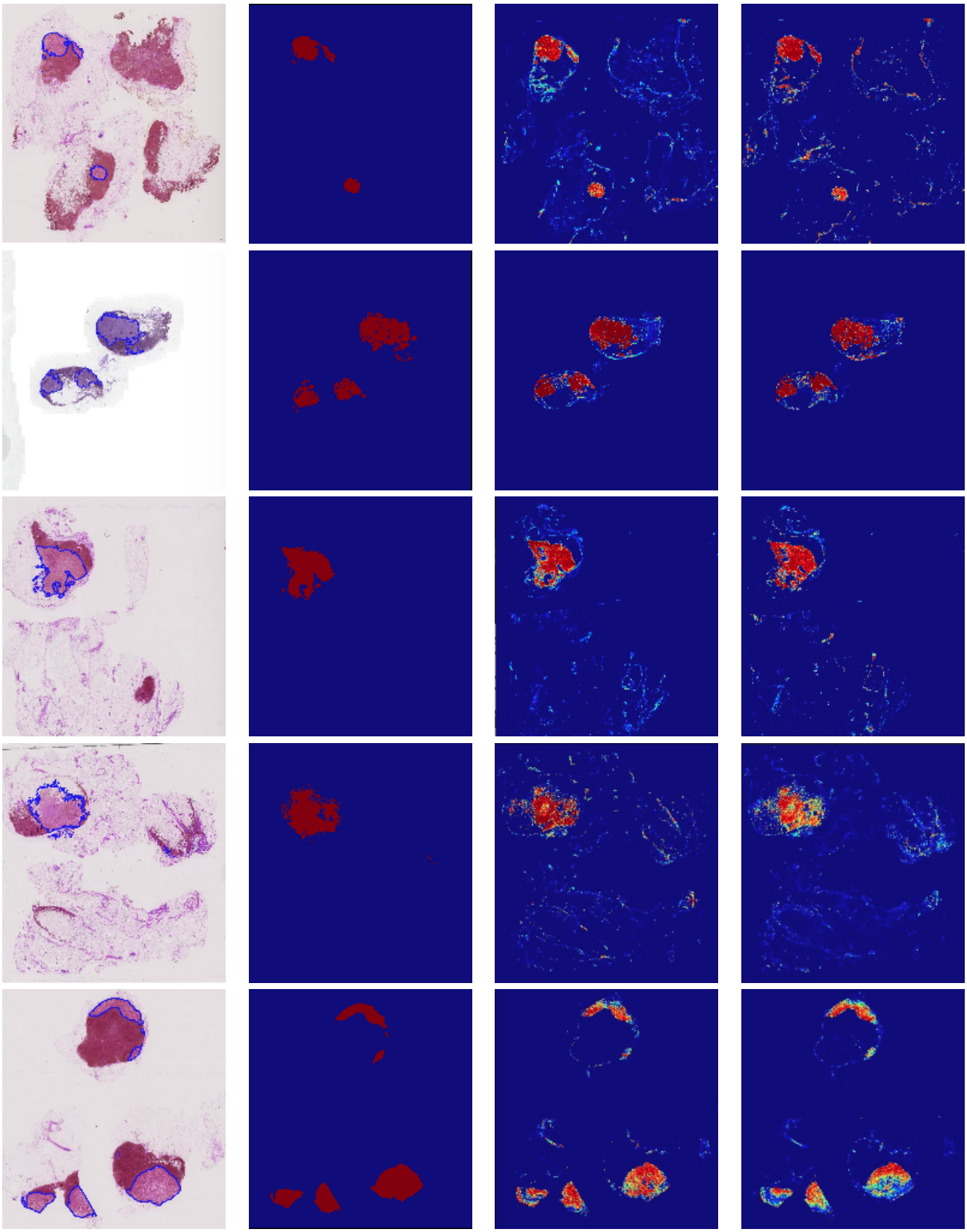


Fig. 11: Some tumor segmentation results of Inception. The first column shows the downsampled pathological images and the area contoured by the blue line is a tumor cell. GT is the ground truth of cancer metastases labeled by the pathologist. The third column shows examples of our detect results. Baseline is the results of training from scratch.

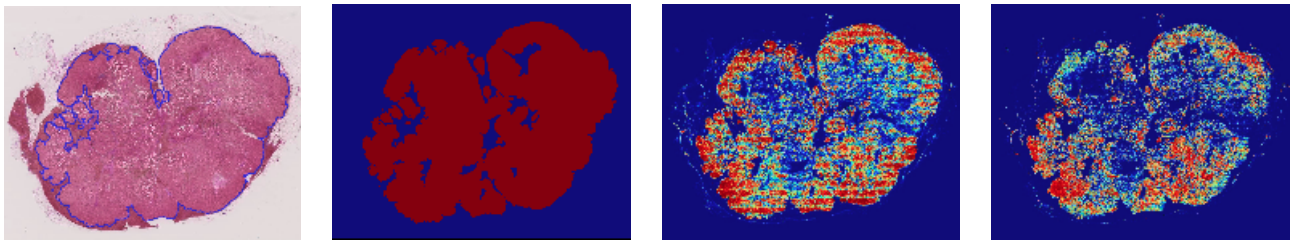


Fig. 12: Failure cases. Few-shot learning still suffers from lack of training data.

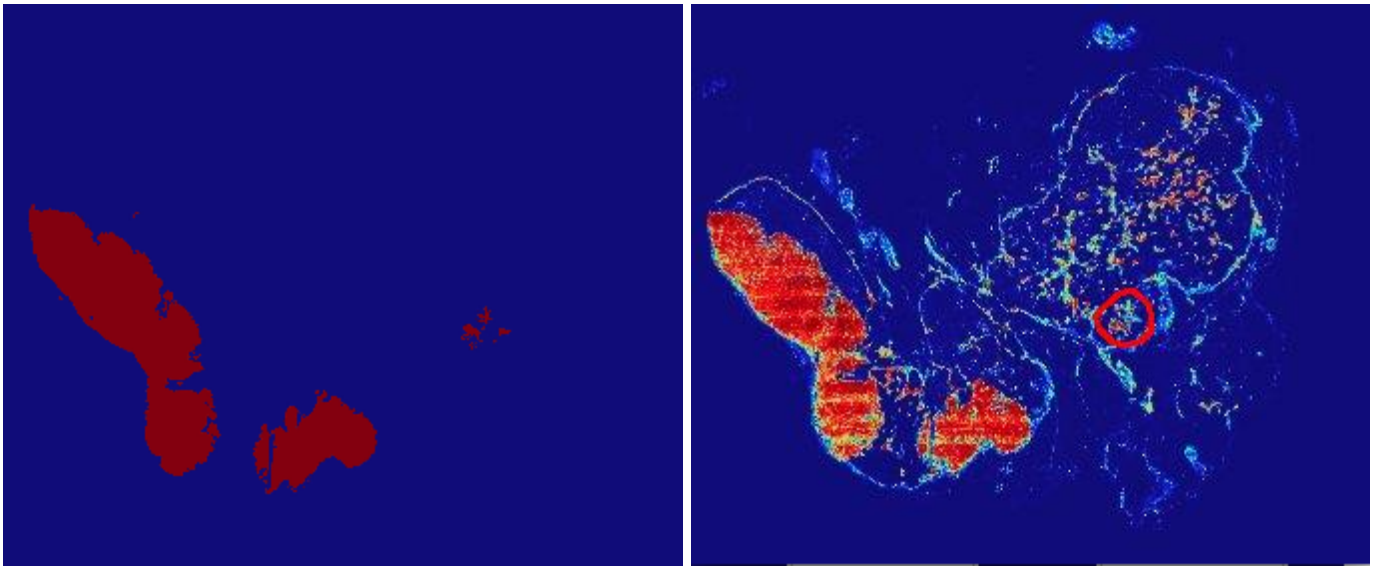


Fig. 13: Ground truth (left) and our detection example (right). The result shows that those small metastases areas are difficult to detect.

information and achieves a better result compared to the supervised method. The proposed method only require slide-level annotations thus we believe it will also benefits those tasks that labeled data is difficult to acquire.

REFERENCES

[1] T. Schindewolf, W. Stolz, R. Albert, W. Abmayr, and H. Harms, "Classification of melanocytic lesions with color and texture analysis using digital image processing," *AQCH*, 1993. 1

[2] A. Esteva, B. Kuprel, R. A. Novoa, J. Ko, S. M. Swetter, H. M. Blau, and S. Thrun, "Dermatologist-level classification of skin cancer with deep neural networks," *Nature*, 2017. 1

[3] T. Wan, J. Cao, J. Chen, and Z. Qin, "Automated grading of breast cancer histopathology using cascaded ensemble with combination of multi-level image features," *Neurocomputing*, 2017. 1

[4] D. C. Cireşan, A. Giusti, L. M. Gambardella, and J. Schmidhuber, "Mitosis detection in breast cancer histology images with deep neural networks," in *MICCAI*, 2013. 1, 2

[5] W. H. Clark, Jr., D. E. Elder, D. Guerry, IV, L. E. Braitman, B. J. Trock, D. Schultz, M. Synnestvedt, and A. C. Halpern, "Model predicting survival in stage i melanoma based on tumor progression," *JNCI*, 1989. 1

[6] S. S. Garapati, L. M. Hadjiiski, K. H. Cha, H.-P. Chan, E. M. Caoili, R. H. Cohan, A. Z. Weizer, A. Alva, C. Paramagul, J. Wei, and C. Zhou, "Urinary bladder cancer staging in ct urography using machine learning." *Medical physics*, 2017. 1

[7] B. Lee and K. Paeng, "A robust and effective approach towards accurate metastasis detection and pn-stage classification in breast cancer," in *MICCAI*, 2018. 1

[8] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *CVPR*, 2015. 1

[9] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei, "Imagenet large scale visual recognition challenge," *IJCV*, 2015. 1

[10] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *CVPR*, 2016. 1

[11] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *CVPR*, 2015. 1

[12] A. Morar, F. Moldoveanu, and E. Grller, "Image segmentation based on active contours without edges," in *ICCP*, 2012. 1

[13] N. R. Pal and S. K. Pal, "A review on image segmentation techniques," *Pattern Recognition*, 1993. 1

[14] B. H. Aksebzeci and . Kayaalti, "Computer-aided classification of breast cancer histopathological images," in *TIPTEKNO*, 2017. 1, 2

[15] N. Habibzadeh Motlagh, M. Jannesary, H. Aboulkheyr, P. Khosravi, O. Elemento, M. Totonchi, and I. Hajirasouliha, "Breast cancer histopathological image classification: A deep learning approach," *bioRxiv*, 2018. 1, 2

[16] F. A. Spanhol, L. S. Oliveira, C. Petitjean, and L. Heutte, "Breast cancer histopathological image classification using convolutional neural networks," in *IJCNN*, 2016. 1, 2

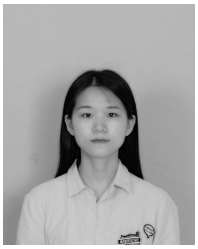
[17] A. Sreekumari, K. S. Shriram, and V. Vaidya, "Breast lesion detection and characterization with 3d features," in *EMBC*, 2016. 1

[18] T. Wollmann and K. Rohr, "Automatic breast cancer grading in lymph nodes using a deep neural network," *CoRR*, 2017. 1, 2

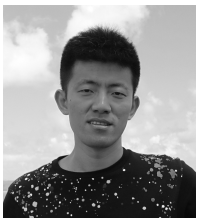
[19] J. Chang, J. Yu, T. Han, H. Chang, and E. Park, "A method for classifying medical images using transfer learning: A pilot study on histopathology of breast cancer," in *Healthcom*, 2017. 2

[20] R. K. Samala, H. Chan, L. Hadjiiski, M. A. Helvie, C. D. Richter, and K. H. Cha, "Breast cancer diagnosis in digital breast tomosynthesis:

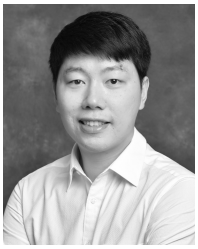
Effects of training sample size on multi-stage transfer learning using deep neural nets," *IEEE Transactions on Medical Imaging*, 2018. 2
[21] N. Bayramoglu, J. Kannala, and J. Heikkil, "Deep learning for magnification independent breast cancer histopathology image classification," in *ICPR*, 2016. 2
[22] S. Han, J. Yang, H. Wan, and J. Yang, "An automated wide-view imaging system of pathological tissue under optical microscopy," in *2nd IET International Conference on Biomedical Image and Signal Processing (ICBISP 2017)*, 2017. 2
[23] M. Amrane, S. Oukid, I. Gagaoua, and T. Ensar, "Breast cancer classification using machine learning," in *EBBT*, 2018. 2
[24] W. Luo, L. Wang, and J. Sun, "Feature selection for cancer classification based on support vector machine," in *GCIS*, 2009. 2
[25] Y. Liu, K. Gadepalli, M. Norouzi, G. E. Dahl, T. Kohlberger, A. Boyko, S. Venugopalan, A. Timofeev, P. Q. Nelson, G. S. Corrado, J. D. Hipp, L. Peng, and M. C. Stumpe, "Detecting cancer metastases on gigapixel pathology images," *CoRR*, 2017. 2
[26] Z. Vang, Yeeleng S.and Chen and X. Xie, "Deep learning framework for multi-class breast cancer histology image classification," in *Image Analysis and Recognition*, 2018. 2
[27] G. Aresta, T. Araújo, S. Kwok, S. S. Chennamsetty, M. S. K. P., A. Varghese, B. Marami, M. Prastawa, M. Chan, M. J. Donovan, G. Fernandez, J. Zeineh, M. Kohl, C. Walz, F. Ludwig, S. Braunewell, M. Baust, Q. D. Vu, M. N. N. To, E. Kim, J. T. Kwak, S. Galal, V. Sanchez-Freire, N. Brancati, M. Frucci, D. Riccio, Y. Wang, L. Sun, K. Ma, J. Fang, I. Koné, L. Boulmane, A. Campilho, C. Eloy, A. Polónia, and P. Aguiar, "BACH: grand challenge on breast cancer histology images," *CoRR*, 2018. 2
[28] H. Lin, H. Chen, S. Graham, Q. Dou, N. Rajpoot, and P. Heng, "Fast scannet: Fast and dense analysis of multi-gigapixel whole-slide images for cancer metastasis detection," *IEEE Transactions on Medical Imaging*, 2019. 2
[29] E. B. B, V. M, J. van Diest P, and et al, "Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer," *JAMA*, 2017. 2, 6
[30] D. Wang, A. Khosla, R. Gargeya, H. Irshad, and A. H. Beck, "Deep learning for identifying metastatic breast cancer," *arXiv preprint arXiv:1606.05718*, 2016. 2, 6, 7
[31] P. Bndi, O. Geessink, Q. Manson, M. van Dijk, M. Balkenhol, M. Hermsen, B. E. Bejnordi, B. Lee, K. Paeng, A. Zhong, Q. Li, F. G. Zanjani, S. Zinger, K. Fukuta, D. Komura, V. Ovtcharov, S. Cheng, S. Zeng, J. Thagaard, A. B. Dahl, H. Lin, H. Chen, L. Jacobsson, M. Hedlund, M. etin, E. Halc, H. Jackson, R. Chen, F. Both, J. Franke, H. Ksters-Vandevelde, W. Vreuls, P. Bult, B. van Ginneken, J. van der Laak, and G. Litjens, "From detection of individual metastases to classification of lymph node status at the patient level: the camelyon17 challenge," *IEEE Transactions on Medical Imaging*, 2018. 3
[32] Q. Huang, F. Zhang, and X. Li, "Few-shot decision tree for diagnosis of ultrasound breast tumor using bi-rads features," *Multimedia Tools and Applications*, 2018. 3
[33] A. Oliver, X. Llad, A. Torrent, and J. Mart, "One-shot segmentation of breast, pectoral muscle, and background in digitised mammograms," in *ICIP*, 2014. 3
[34] N. Zemmal, N. Azizi, N. Dey, and M. Sellami, "Adaptive semi supervised support vector machine semi supervised learning with features cooperation for breast cancer classification," *Journal of Medical Imaging and Health Informatics*, 2016. 3
[35] W. Sun, B. Tseng, J. Zhang, and W. Qian, "Enhancing deep convolutional neural network scheme for breast cancer diagnosis with unlabeled data," *Computerized Medical Imaging and Graphics*, 2016. 3
[36] J.-B. Li, Y. Yu, Z.-M. Yang, and L.-L. Tang, "Breast tissue image classification based on semi-supervised locality discriminant projection with kernels," *Journal of Medical Systems*, 2012. 3
[37] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," *CoRR*, 2015. 4
[38] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *CoRR*, 2014. 4
[39] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," *CoRR*, 2015. 4



Jiaojiao Chen is a M.Sc. student in the School of Computer Science and Engineering, South China University of Technology. She obtained my B.Sc. degree from Hubei University. Her research interests include deep learning and medical imaging.



Jianbo Jiao is a Postdoctoral Researcher in the Department of Engineering Science at the University of Oxford, and a member of the Biomedical Image Analysis (BioMedIA) group and the Visual Geometry Group (VGG). He obtained his Ph.D. in Computer Science from City University of Hong Kong. His research interests include computer vision and machine learning.



Shengfeng He is an Associate Professor in the School of Computer Science and Engineering, South China University of Technology. He was a Research Fellow at City University of Hong Kong. He obtained his B.Sc. degree and M.Sc. degree from Macau University of Science and Technology, and the Ph.D degree from City University of Hong Kong. His research interests include computer vision, image processing, computer graphics, and deep learning.

Guoqiang Han received the B.Sc. degree from the Zhejiang University, Hangzhou, China, in 1982, and the masters and Ph.D. degrees from the Sun Yat-sen University, Guangzhou, China, in 1985 and 1988, respectively. He is a Professor with the School of Computer Science and Engineering, South China University of Technology, Guangzhou. He was the dean of the School of Computer Science and Engineering. He has published over 100 research papers. His current research interests include multimedia, computational intelligence, machine learning, and computer graphics.

Jing Qin received his PhD degree in Computer Science and Engineering from the Chinese University of Hong Kong in 2009. He has been an assistant professor at The Hong Kong Polytechnic University from 2016. His research interests include visualization, human-computer interaction and deep learning.