MSA-Net: Multiscale Spatial Attention Network for the Classification of Breast Histology Images

Zhanbo Yang[1,2], Lingyan Ran[2], Yong Xia[1,2(✉)], and Yanning Zhang[2]

[1] Research & Development Institute of Northwestern Polytechnical University in Shenzhen, Shenzhen 518057, China
yxia@nwpu.edu.cn

[2] National Engineering Laboratory for Integrated Aero-Space-Ground-Ocean Big Data Application Technology, School of Computer Science and Engineering, Northwestern Polytechnical University, Xi'an 710072, China

Abstract. Breast histology images classification is a time- and labor-intensive task due to the complicated structural and textural information contained. Recent deep learning-based methods are less accurate due to the ignorance of the interfering multiscale contextual information in histology images. In this paper, we propose the multiscale spatial attention network (MSA-Net) to deal with these challenges. We first perform adaptive spatial transformation on histology microscopy images at multiple scales using a spatial attention (SA) module to make the model focus on discriminative content. Then we employ a classification network to categorize the transformed images and use the ensemble of the predictions obtained at multiple scales as the classification result. We evaluated our MSA-Net against four state-of-the-art methods on the BACH challenge dataset. Our results show that the proposed MSA-Net achieves a higher accuracy than the rest methods in the five-fold cross validation on training data, and reaches the 2nd place in the online verification.

Keywords: Breast cancer · Histology image classification · Multiscale · Spatial attention · Convolutional neural networks.

1 Introduction

Breast cancer is one of the severe types of cancers in women, which accounts for 25.16% of all cancers with 1.68 million new cases worldwide in 2012 [11]. During the diagnosis of breast cancer, hematoxylin-eosin (H&E) stained histology images of tissue regions resulted from needle biopsy are evaluated to determine the type, including normal, benign, in situ carcinoma, and invasive carcinoma. Due to the complexity of histology images, detecting carcinoma by pathologists is time-consuming, labor-intensive, and subjective. The scientific community has been working on the development of automated detection and diagnosis tools over the past years. For instance, the Grand Challenge on BreAst Cancer Histology images (BACH) [2] organized in conjunction with the 15th International

Conference on Image Analysis and Recognition (ICIAR 2018) aims at the classification and segmentation of H&E stained breast histology microscopy images.

Automated classification of H&E stained breast histology microscopy images is challenging in two aspects. First, microscopy images usually have an extremely high resolution, and hence contain rich structural information and details, which are hard to be characterized effectively at a single scale. Second, microscopy images from different categories may exhibit partly overlapped patterns, which interfere carcinoma detection, such as the hard mimics from benign lesion which have similar morphological appearance with carcinoma.

To address both issues, various deep learning-based methods have been designed as a result of the success of deep convolutional neural networks (DCNNs) in computer vision [4, 1, 3, 10, 9, 13, 14, 5]. Araujo et al. [1] proposed a patches-based DCNN + SVM model to address the breast microscopy image classification problem. In this model, a DCNN is designed for feature extraction and a support vector machine (SVM) is used as a classifier. Chennamsetty et al. [3] constructed an ensemble of three DCNNs, each of which was pre-trained on different pre-processing regimes, and achieved the 1st place on the BACH challenge at the first stage. Besides, attention-based methods [7] were also proposed for this purpose. For instance, following the design trends of squeeze-and-excitation network (SE-Net) [7], Vu et al. [12] incorporated the self-attention mechanism into an encoder-decoder network. Despite the improved performance, these DCNN-based methods still suffer from less-discriminative power resulted mainly from the inadequate quantity of training data. We suggest exploring the multiscale and spatial attention aided contextual information, which have been commonly used by human histology image reader.

In this paper, we propose the multi-scale spatial attention deep convolutional neural network (MSA-Net) for the automated classification of H&E stained breast histology microscopy images. To exploit the multiscale information of images, we first convert each image to three scales, then perform adaptive spatial transformation on the microscopy patches cropped at each scale by the spatial attention (SA) module, which is followed by a classification network to categorize the transformed patches, and finally combine the results to generate the image label. We expect that can learn how to perform spatial transformation on the microscopy patches for precise classification. We evaluated the proposed algorithm on the BACH challenge dataset and achieved an accuracy of $94.50 \pm 1.27\%$ in the five-fold cross validation on training data and an accuracy of 94.00% in the online verification.

2 Method

Given a H&E stained breast histology microscopy image $X \in \mathbb{R}^{H \times W \times C}$, our goal is to predict the image label $Y \in \{0, 1, 2, 3\}$, which includes four classes: Normal (0), Benign (1), In situ carcinoma (2), and Invasive carcinoma (3). The proposed MSA-Net algorithm consists of three steps: (1) multiscale image patch extraction, (2) SA-Net based image patch classification, and (3) multi-branch

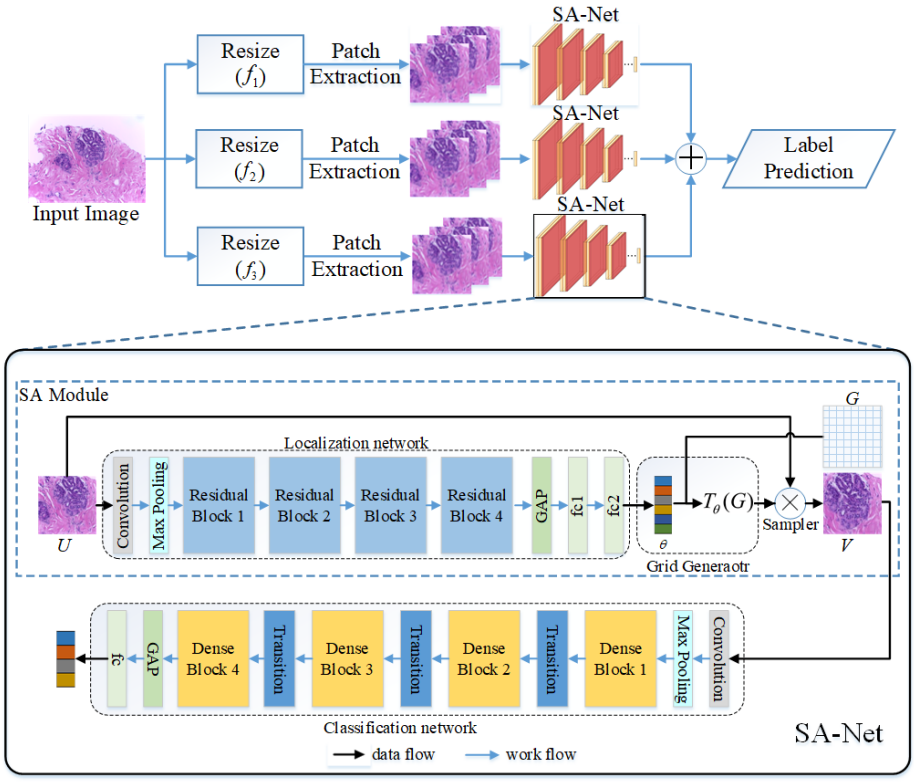


Fig. 1. Diagram of the proposed MSA-Net algorithm. For a histology microscopy image, we first extract microscopy patches at multiple scales, then classify these patches by SA-Net, and finally predict image label by ensemble of the classification results. The SA-Net includes two parts: the SA module consisting of localization network, grid generator and sampler, and the classification network. An input patch U is passed to localization network which regresses the transformation parameters θ, then the regular spatial grid G over V is transformed to the sampling grid $T_\theta(G)$, which is applied to U, and producing the warped output patch V, and lastly V is passed to classification network to get label prediction.

ensemble. The diagram that summarizes this algorithm is shown in Fig. 1. We now delve into the details of each step.

2.1 Multiscale image patch extraction

For an breast histology microscopy image with size of $H \times W$, we first down-sample the images by factors f_1, f_2 and f_3 to get resized images at three scales, where the down-sampling factor $f \in [1, inf)$ with $f = 1$ being the original image. Then we slide a $h \times w$ window with a stride of s on the resized images at each scale for extracting multiscale microscopy patches. In this way, the number of

microscopy patches we extracted from an image is

$$N = \left(\left\lfloor \frac{W/f - w}{s} \right\rfloor + 1 \right) \times \left(\left\lfloor \frac{H/f - h}{s} \right\rfloor + 1 \right) \tag{1}$$

where f is the down-sampling factor, and $\lfloor \rfloor$ denotes rounding down. Note that the resized images were divided into partly overlapped patches to generate more training data. Next, the intensities of each microscopy patch are standardized to zero mean and unit variance.

To alleviate overfitting of SA-Net, we employ two data augmentation methods to increase the diversity of the training dataset. First, each microscopy patch is augmented into eight patches by rotating an angle of $k \cdot \pi/2$, where $k = \{0, 1, 2, 3\}$, and with/without vertical reflection. Second, random color perturbations have been applied to each patch.

2.2 SA-Net based image patch classification

The proposed SA-Net including two parts: the SA module for performing adaptive spatial transformation on inputs, and the classification network for predicting the label of transformed patches.

Spatial attention module. Due to the inter- and intra-confusing structural and textural information, we perform adaptive spatial transformation on microscopy patches by SA module before categorizing them. The SA module is split into three parts: (i) localization network, (ii) grid generator, and (iii) sampler, as shown in Fig. 1.

First, a localization network takes the extracted microscopy patch $U \in \mathbb{R}^{h \times w \times C}$, with h, w and C being the height, width and number of channels respectively, and outputs the transformation parameters θ. Due to their outstanding performance in non-linear transformation, we choose residual network [6] with 152 learnable layers named ResNet-152 as the backbone network for the localization network. ResNet-152 includes a convolutional layer with the kernel size of 7×7, a 3×3 max pooling layer, four residual blocks, which have 3, 8, 36, and 3 triple-layer residual groups, respectively, and an average pooling layer followed by the softmax operation. To adapt to our problem, we remove the classification layer and add two weight layers to predict the transformations: (i) fully-connected layer to reduce the length of feature vectors from 1024 to 128; (ii) fully-connected layer with 6-D output.

Then, the predicted transformation parameters are used to create a sampling grid by the grid generator, which is a set of points where the input map should be sampled to produce the transformed patch $V \in \mathbb{R}^{h \times w \times C}$. In detail, the output pixels are defined to lie on a regular grid $G = \{G_i\}$ of pixels $G_i = (x_i^O, y_i^O)$, where $i \in [0, 1, \ldots, hw - 1]$, forming an output patch V. The spatial transformation formula is

$$\begin{pmatrix} x_i^I \\ y_i^I \end{pmatrix} = \mathrm{T}_\theta(G_i) = A_\theta \begin{pmatrix} x_i^O \\ y_i^O \\ 1 \end{pmatrix} = \begin{bmatrix} \theta_{11} & \theta_{12} & \theta_{13} \\ \theta_{21} & \theta_{22} & \theta_{23} \end{bmatrix} \begin{pmatrix} x_i^O \\ y_i^O \\ 1 \end{pmatrix} \tag{2}$$

where (x_i^O, y_i^O) are the target coordinates of the regular grid in the output patch, (x_i^I, y_i^I) are the source coordinates in the input patch that define the sample points, and A_θ is a 6-DoF affine transformation matrix.

Lastly, a sampler takes the set of sampling points, along with the input patch U and produces the sampled output patch V. For doing that, bilinear interpolation sampling is applied in coordinates define the spatial location in U to generate the value at a particular pixel in V, where bilinear interpolation sampling is an extension of linear interpolation sampling for interpolation function of two variables on a rectilinear 2D grid.

Classification network. For the transformed microscopy patches, we categorize them by classification network which using fine-tuned DenseNet-161 [8] model. Similar to ResNet models, DenseNet-161 consists of 161 learnable layers, including a convolutional layer with the kernel size of 7×7, a 3×3 max pooling layer, four dense blocks, three transition layers, and a global average pooling layer followed by the softmax operation. Those four dense blocks contain 6, 12, 36, and 24 dual-layer dense groups, respectively. To adapt the DenseNet-161 to our problem, we keep only four neurons in the layer before softmax.

2.3 Ensemble of the classification results

Through the SA-Net, a microscopy patch is spatially transformed and recognized as one of the four classes. The probabilities that a resized image belongs to a category are determined by the ratio of number of patches belongs to this category to the number of patches extracted from this image. Finally, the category of each histology microscopy image is recognized by the average prediction on three images resized from that image.

2.4 Training procedure

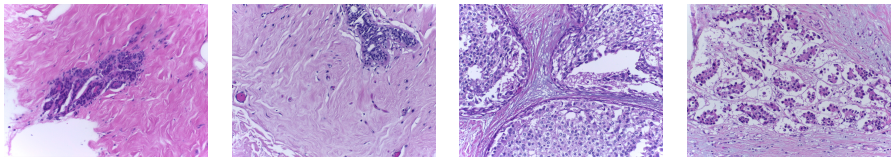
Since each module of MSA-Net is differentiable, we train this deep learning model in an end-to-end fashion. To avoid the impact of the SA module in the initial training stage, the final layer of localization net is initialized to regress the identity transform of input patches. To train the SA-Net at each scale, we adopt the Adam optimizer to minimize the cross entropy loss, and set the batch size to 16, learning rate to 0.0001 with a decay of 10% every 10 training epochs, and the maximum epoch number to 30.

3 Experiments and Results

3.1 BACH dataset

The ICIAR 2018 grand challenge on BACH dataset [2] was used for this study. This dataset is composed of 400 H&E stained breast histology microscopy images with a size of 2048×1536 for training a classification model and 100 similar

images with the same size for testing. Training images have four class labels including normal, benign, in situ carcinoma and invasive carcinoma (see Fig. 2). Each of four category contains 100 training images. The 100 testing images were officially presented for online verification, and their labels are not available.



(a) Normal (b) Benign (c) In situ (d) Invasive

Fig. 2. Four H&E stained breast histology microscopy images from different categories.

3.2 Implementation details

During offline training procedure, we evaluated the proposed MSA-Net algorithm using 400 training images with the five-fold cross validation.

In the training stage, microscopy patches of size 224×224 were extracted from 320 training images at three scales and were augmented to train SA-Net. We set the down-sampling factor f_1 as 2 which means the size of resized images is 1024×768(Scale I), and f_2 as 4 which means the size of resized images is 512×384(Scale II). The minimum size of image we down-sampled is 296×224(Scale III), which corresponds to f_3 as 6.86. We set the stride to 133 at Scale I, 36 at Scale II, and 5 at Scale III for training patch extraction. In the testing stage, the patches were extracted from resized testing images with twice strides against training data for computing acceleration.

3.3 Results

We show the accuracy, precision, recall, the area under the receiver operating characteristic (ROC) curve (AUC), and the ROC curve of the proposed algorithm in differentiating each category of images and the overall classification accuracy in Table 1 and Fig. 3. It shows that it is easy for the proposed algorithm to separate invasive carcinoma tissues from others, but difficult to separate normal tissues from others. Nevertheless, we achieved an average AUC of more than 98.26% in all categories and an overall accuracy of $94.50 \pm 1.27\%$, which demonstrate the effectiveness of the proposed algorithm in the classification of breast histology microscopy images.

Next, we compare the proposed MSA-Net algorithm to four recent methods: (1) using a custom DCNN for feature extraction and a SVM for classification [1], (2) using a pre-trained VGG-16 together with a variety of data augmentation methods [2], (3) using a pre-trained Inception-Resnet-v2 together with a training

Table 1. Performance (mean ± standard deviation) of the proposed MSA-Net algorithm on BACH training images with five-fold cross validation.

Category	Accuracy(%)	Precision(%)	Recall(%)	AUC(%)
Normal	96.50±1.84	93.04±3.91	93.00±4.00	99.57±0.24
Benign	97.00±1.00	95.02±2.90	93.00±4.00	99.61±0.26
In situ	97.25±2.15	93.10±7.45	97.00±2.45	99.10±0.79
Invasive	98.25±1.87	98.05±2.39	95.00±4.75	98.26±1.29
All	**94.50±1.27**	94.80±4.16	94.50±3.80	99.14±0.65

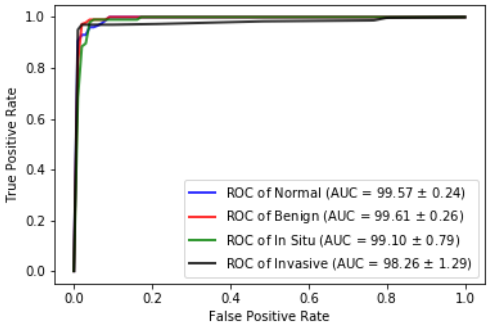


Fig. 3. The ROC curves. The True positive Rate and False Positive Rate are calculated through a one-vs.-rest strategy based on the classification results.

process of two stages [2], and (4) using ensemble of three pre-trained DCNNs [2]. Table 2 shows the overall accuracy of those four methods, and the accuracy of our algorithm. It reveals that proposed algorithm is substantially more accurate than those four methods on this image classification task.

Table 2. Accuracy (%) of the proposed MSA-Net algorithm on BACH training dataset using five-fold cross valiation and four leading methods.

Method	Accuracy(%)
DCNN+SVM,2017 [1]	77.80
Pre-trained VGG-16,2018 [2]	83.00
Pre-trained Inception-Resnet-v2, 2018 [2]	87.00
Ensemble of three DCNNs, 2018 [2]	87.00
MSA-Net (proposed)	**94.50**

Moreover, besides 400 labeled training images, the organizers of BACH challenge also provided 100 microscopy images without labels for online testing. We submitted the classification results, which we obtained on those testing images, to the official website, and the organizers calculated the accuracy of our algo-

rithm. We synthesized the leader-board of the challenge at first and second stages and displayed it in Table 3. It shows that our algorithm achieved an accuracy of 94.00% in the online validation and won the 2nd place on the leader-board.[3]

Table 3. The leader-board of the BACH challenge on testing dataset (ranked by accuracy on Task Part A).

Position	Participants	Accuracy(%)
1	hanwang.0501	95.00
2	young(ours)	**94.00**
3	bamboo	93.00
4	HeechanYang	93.00
5	YUN1503	92.00
6	heechan	92.00

To further demonstrate the validity of the proposed MSA-Net, we also design the experiment on another histopathological dataset named Breast Cancer Histopathological Database (BreakHis), which is composed of 2,480 benign and 5,429 malignant microscopic images of breast tumor tissue with 700×460 pixels and 3 channels collected from 82 patients using different magnifying factors (40X, 100X, 200X, and 400X). To match the resolution of images in BACH dataset, we use microscopic images with 40X magnifying factor for experimentation, which contains 625 benign and 1370 malignant samples. We randomly select 20% of microscopic images for testing and other images for training. As shown in Table 4, we achieved an AUC of 99.99% in all categories and an overall accuracy of 99.75%, which further demonstrate the effectiveness of the proposed algorithm in the classification of breast microscopy images.

Table 4. Performance of the proposed MSA-Net algorithm on BreakHis dataset.

Category	Accuracy(%)	Precision(%)	Recall(%)	AUC(%)
Benign	99.75	100.00	99.20	99.99
Malignant	99.75	99.65	100.00	99.99
All	**99.75**	99.83	99.60	99.99

4 Discussion

4.1 Choice of down-sampling factor

The down-sampling factor f represents the resolution of histology microscopy images input to the SSA-Net, hence plays an important role in classifying images.

[3] Available at: https://iciar2018-challenge.grand-challenge.org/evaluation/results/

To determine the best value of f, we performed the proposed algorithm with different values of f and show their accuracy in Table 5. Due to the size of patches we input SA-Net are 224×224, the minimum size of image we down-sampled is 296×224, which corresponds f as 6.86. Table 5 shows that when f is 4, the proposed algorithm has best accuracy, however, when f is 1, the algorithm has worst accuracy. Hence we empirically set f to 2, 4 and 6.86 respectively in our experiments.

Table 5. Accuracy of the proposed algorithm with different values of down-sampling factor f.

f	Size of resized image	Accuracy(%)
1	2048×1536	86.25
2	1024×768	92.50
4	512×384	93.75
6.86	296×224	91.25

4.2 Time complexity

The experiments were performed on a PC (Intel Core i7-4790 CPU 3.2GHz, NVidia GTX 1080Ti GPU and 32GB memory) with Ubuntu 14.04 64bit system. It took about 63.5 hours in the training stage and less than 5 seconds in the testing stage when applying the proposed algorithm to classify each microscopy image. Despite the fact that training the model is time-consuming, our approach can be suitably fitted to a routine clinical workflow with pretty fast testing efficiency.

5 Conclusion

In this paper, we propose the MSA-Net algorithm to classify H&E stained breast microscopy images into four categories including normal, benign, in situ carcinoma, invasive carcinoma. Our results demonstrate the superior performance of the proposed algorithm with the 2nd place on the BACH challenge official leaderboard and a five-fold cross validation accuracy of $94.50 \pm 1.27\%$ on BACH training images. In the future, the proposed MSA-Net algorithm serves great potential to the development of semi-supervision mechanism when identifying microscopy images using unlabeled samples.

Acknowledgement. This work was supported in part by the Science and Technology Innovation Committee of Shenzhen Municipality, China, under Grant JCYJ20180306171334997, in part by the National Natural Science Foundation of China under Grant 61771397 and 61902322, in part by the Fundamental Research Funds for the Central Universities under Grant 3102019G2019KY0001,

in part by the Seed Foundation of Innovation and Creation for Graduate Students in Northwestern Polytechnical University under Grants ZZ2019029, and in part by the Project for Graduate Innovation team of Northwestern Polytechnical University.

References

1. Araújo, T., Aresta, G., Castro, E., Rouco, J., Aguiar, P., Eloy, C., Polónia, A., Campilho, A.: Classification of breast cancer histology images using convolutional neural networks. PlOS ONE **12**(6), 1–14 (2017)
2. Aresta, G., Araújo, T., Kwok, S., Saketh Chennamsetty, S., Safwan, M., Alex, V., Marami, B., Prastawa, M., Chan, M., Donovan, M., et al.: BACH: Grand challenge on breast cancer histology images. arXiv preprint arXiv:1808.04277 (2018)
3. Chennamsetty, S.S., Safwan, M., Alex, V.: Classification of breast cancer histology image using ensemble of pre-trained neural networks. In: Proceeding of the International Conference on Image Analysis and Recognition(ICIAR). pp. 804–811. Springer (2018)
4. Deng, J., Dong, W., Socher, R., Li, L.J., Li, K., Fei-Fei, L.: ImageNet: A large-scale hierarchical image database. In: Proceeding of the IEEE conference on computer vision and pattern recognition(CVPR). pp. 248–255. IEEE (2009)
5. Han, J., Zhang, D., Cheng, G., Guo, L., Ren, J.: Object detection in optical remote sensing images based on weakly supervised learning and high-level feature learning. IEEE Transactions on Geoscience and Remote Sensing **53**(6), 3325–3337 (2014)
6. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceeding of the IEEE conference on computer vision and pattern recognition(CVPR). pp. 770–778 (2016)
7. Hu, J., Shen, L., Sun, G.: Squeeze-and-excitation networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition(CVPR). pp. 7132–7141 (2018)
8. Huang, G., Liu, Z., Van Der Maaten, L., Weinberger, K.Q.: Densely connected convolutional networks. In: Proceeding of the IEEE conference on computer vision and pattern recognition(CVPR). pp. 2261–2269. IEEE (2017)
9. Ragab, D.A., Sharkas, M., Marshall, S., Ren, J.: Breast cancer detection using deep convolutional neural networks and support vector machines. PeerJ **7**, e6201 (2019)
10. Ren, J.: Ann vs. svm: Which one performs better in classification of mccs in mammogram imaging. Knowledge-Based Systems **26**, 144–153 (2012)
11. Stewart, B., Wild, C.P., et al.: World cancer report 2014. The International Agency for Research on Cancer (2014)
12. Vu, Q.D., To, M.N.N., Kim, E., Kwak, J.T.: Micro and macro breast histology image analysis by partial network re-use. In: Proceeding of the International Conference on Image Analysis and Recognition(ICIAR). pp. 895–902. Springer (2018)
13. Wang, Z., Ren, J., Zhang, D., Sun, M., Jiang, J.: A deep-learning based feature hybrid framework for spatiotemporal saliency detection inside videos. Neurocomputing **287**, 68–83 (2018)
14. Yan, Y., Ren, J., Sun, G., Zhao, H., Han, J., Li, X., Marshall, S., Zhan, J.: Unsupervised image saliency detection with gestalt-laws guided optimization and visual attention based refinement. Pattern Recognition **79**, 65–78 (2018)