MULTI-CLASSIFICATION OF BREAST CANCER HISTOLOGY IMAGES BY USING GRAVITATION LOSS

Zhu Meng[⋆] *Zhicheng Zhao*[⋆†] *Fei Su*[⋆†]

[⋆] Beijing University of Posts and Telecommunications, China
[†] Beijing Key Laboratory of Network System and Network Culture, China

ABSTRACT

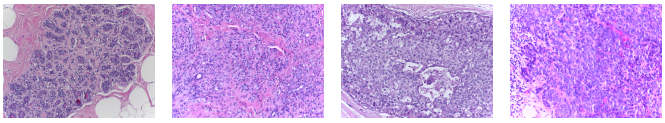
The scarcity of professional doctors stimulates the progress of breast cancer classification. However, there are still numerous challenges such as varied appearances (color, texture etc.) of microscopy images and the ambiguous category boundaries. In this paper, we propose an efficient and effective method to achieve multi-classification for H&E stained breast cancer images. Firstly, to restrain color noises in the staining stage, data augmentation in HSV color space is used to increase the diversity of color distribution. In addition, inspired by the principle of gravitation, a Gravitation Loss (G-loss) is proposed to maximize inter-class difference and minimize intra-class variance. The experimental results on public BACH 2018 dataset indicate that the proposed algorithm achieves the state-of-the-art performance, which demonstrates its effectiveness.

***Index Terms*—** Multi-classification, Breast Cancer, Histology Images, Gravitation Loss, Deep Learning

1. INTRODUCTION

Until 12 September 2018, the number of breast cancer diagnoses has reached approximately 2.1 million, which contributes 11.6% of the total cancer incidence [1]. Despite the rapid development of medical technology, breast cancer diagnosis is still a difficult and exhausting task. The insufficiency of professional doctors and the complexity of pathology images has intensified this status, especially in developing and under-developing countries. Therefore, a computer-aided cancer diagnosis has a potential application value.

Numerous methods [2][3][4][5] were proposed to deal with breast cancer classification, and Hematoxylin and Eosin (H&E) stained histology images were commonly used. For example, both Camelyon16 challenge [6] and ICIAR 2018 Grand Challenge on BreAst Cancer Histology Images (BACH 2018) [7] adopted H&E stained images. However, due to the differences in lab protocols, concentration, source manufacturer, scanners, and even staining time, the results of H&E stained images are also various [8]. Fig.1 shows four categories from BACH 2018, in which different color distributions are obvious. Color uncertainty increases the difficulty of automatic diagnosis. To alleviate the impacts of color variability, instead of forcing all images to be similar, we augment the training data by a color jitter operation in HSV color space.



(a) (b) (c) (d)

Fig. 1: Microscopy images from BACH 2018 are of four classes: (a) normal, (b) benign, (c) in situ carcinoma and (d) invasive carcinoma.

Images in BACH 2018 are labelled as four classes, i.e., normal, benign, in situ carcinoma and invasive carcinoma. However, due to the subjective difference and the pathology complexity, different categories usually have ambiguous category margins. In order to optimize the margins of inter-class and intra-class, we propose a Gravitation Loss based on the principle of gravitation. Take the galaxy as an example, mutual gravitation of planets to the Earth in our solar system is much stronger than that of arbitrary planets outside solar system. In analogy with the gravity, we abstract the similarities between images as two kinds of forces: inter-class gravity and intra-class one. Then similar intra-class features and discriminative inter-class features can be learned by Gravitation Loss and the classification performance will be improved.

Extensive experiments on BACH 2018 dataset demonstrate the effectiveness of this method. The contributions can be summarized into three aspects:
(1) A data augmentation method in HSV color space is proposed to improve the color generalization ability in the multi-classification model,
(2) A novel Gravitation Loss is proposed to regulate loss function so as to optimize the inter-class and intra-class margins,
(3) Experiments on BACH 2018 dataset show that our method achieves the state-of-the-art results.

2. RELATED WORK

Many experiments were conducted on Camelyon 16 dataset, while only binary classification results were recognized. In

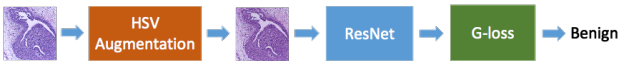


Fig. 2: The whole architecture.

order to meet the clinical requirement, great efforts of multi-classification have been made recently. For instance, based on BACH 2018 dataset, numerous approaches were proposed to classify H&E breast cancer images into four categories [2][3][4][5][9][10].

To obtain more training data and reduce the number of network parameters, most of the previous methods explored patch-based algorithms [2][3][5][10]. However, each microscopy image in the size of 2048×1536 has only one single label, which means that all patches from the same image would be labelled as the same label regardless of whether the patches include the corresponding information. This rough processing will introduce many wrong tags, which will certainly reduce the classification quality. To solve this problem, Xie [11] removed some mislabelled patches via training an independent CNN model first and then training another one for classification. However, it was a time-consuming operation and did not provide superior results. Different from previous approaches, we firstly downsample input images by a 0.25 scale factor to reduce parameters, and then apply random cropping to enhance data diversity, which is efficient and effective.

To deal with the uncertainty of staining, two common methods were applied: (1) normalize the color of all images and make them look similar, (2) increase the model's generalization ability by enhancing color variation. For the former, the methods usually chose a source image first and then tuned the color of other images [2] [3]. Although all the images would be somewhat alike to the source one, the methods largely relied on the choice of the source image. For the latter, Shaban [12] generated many different color images via GAN [13], which was complicated and inconvenient. To simplify the operation, we practice color jitter in HSV color space, which shows a high performance.

Although many loss functions have been proposed for natural image processing [14], it is scarce in medical images. Milletari [15] proposed a Dice Loss, while it was only effective in medical image segmentation. For medical image classification, few loss functions were applied. Moreover, the differences of natural image categories are obvious but ambiguous for breast cancer ones. Different from previous works only focusing on complex normalization and large networks, a succinct Gravitation Loss is proposed in this paper to maximum the inter-class margins and minimize the intra-class variances.

3. METHODOLOGY

The framework of our method is shown in Fig.2. As mentioned before, the stained images may present in various colors. To overcome the impact of color differences in classification, data augmentation on HSV color space is first applied and we will introduce it in Section 3.1. Additionally, to solve the problem that the boundaries of different categories are ambiguous, a Gravitation Loss is proposed to enlarge the margins between classes as well as decrease the variance within the class. It will be described in Section 3.2.

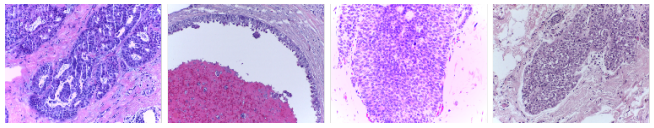
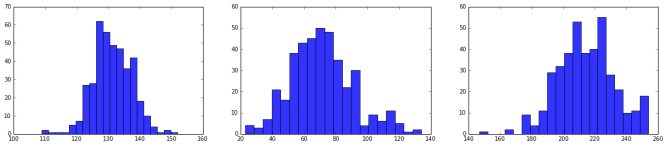


Fig. 3: Several in situ carcinoma H&E stained samples with different color distributions from BACH 2018.



(a) H-channel (b) S-channel (c) V-channel

Fig. 4: The mean distribution of each channel of training samples from BACH 2018.

3.1. HSV Augmentation

The color of H&E stained images is influenced by many factors. Take the microscopy images from BACH 2018 as examples, Fig.3 shows several in situ carcinom images with different color distributions. Since different images in the same class may present various colors, the impact of image color on classification should be limited to some extent. Therefore, the color of the stained sections should be more variegated. Specifically, according to human visual perception for color changes, we apply a data augmentation scheme on uniform HSV color space to emphasize morphological changes. To further prove the difference in color of the stained images, the mean of each channel of every image is calculated in the training set and Fig.4 shows the distributions. According to the distribution of each channel of HSV, a Gaussian color jitter is combined in each channel of the input images. Let x be the original pixel, then the new pixel value x' is defined as

$$x' = clip(x + f(G)) \tag{1}$$

where $f(G)$ denotes Gaussian noise. For H-channel, the values are clipped to $[0, 180]$ and $[0, 255]$ for S- and V-channel. As shown in Fig.5, HSV augmentation increases color changes .

3.2. Gravitation Loss

The boundaries of different categories of breast cancer images are blurry. For example, the images from different classes (shown in Fig.1) are similar to each other. To distinguish images of different classes more clearly, a Gravitation Loss

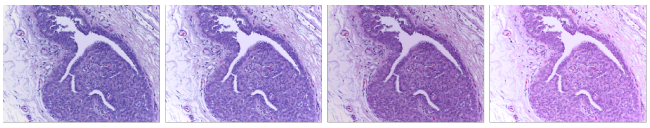


Fig. 5: Examples of color jitter.

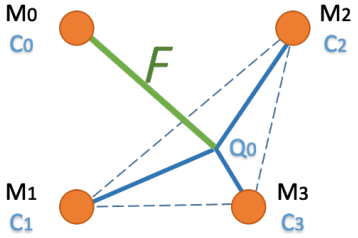


Fig. 6: F denotes gravitation force between class 0 and other classes, where C_i is the center of class i, M_i is the mass of class i ($i \in \{0, 1, 2, 3\}$) and Q_0 is the center of mass of class 1, 2 and 3.

(G-loss) is proposed to increase the inter-class margins while decreasing the intra-class ones.

As Newton's law of universal gravitation described, every point mass attracts every single other point mass by a force acting along the line intersecting both points [16]:

$$F = \frac{Gm_1m_2}{r^2} \tag{2}$$

where F is the gravitation force, G is the gravitational constant, m_1 and m_2 are the masses of the objects and r is the distance between them. Similarly, there is gravitation force between two different classes. Assuming that the mass of each sample is 1, the mass of one class can be denoted as the number of samples in this class. As shown in Fig.6, F denotes gravitation force between class 0 and other classes, M_i and C_i are respectively mass and the center of class i, Q_0 denotes the center of mass of class 1, 2 and 3. Since the gravitation force between different classes is considered the smaller the better, it can be designed as a part of the loss function, i.e., G-loss. Specially, we call this kind of inter-class loss $L_{G-inter}$:

$$L_{G-inter} \propto F \tag{3}$$

Similarly, for a batch during training, the loss for class 0 can be presented as:

$$L^{(0)}_{G-inter} \propto G \cdot \frac{M_0 \left(M_1 + M_2 + M_3\right)}{\|C_0Q_0\|^2} \tag{4}$$

F for other classes is similar to class 0. To be specific, the constant G is ignored and the whole G-loss of different classes is designed as:

$$L_{G-inter} = \frac{1}{4}\sum_{k=0}^{3} L^k_{G-inter} \tag{5}$$

$$L_{G-inter} = \frac{1}{4}\sum_{k=0}^{3} \frac{M_k \left(n - M_k\right)}{\|C_kQ_k\|^2} \tag{6}$$

where n denotes the batch-size.

Like inter-class loss, the intra-class loss can also be defined according to gravitation force. Since the force between samples of the same class is considered the larger the better, the loss will be inversely proportional to the gravitation force:

$$L_{G-intra} \propto \frac{1}{F} \tag{7}$$

Since the mass of each sample S_k is 1, the whole intra-class loss is given as:

$$L_{G-intra} = \frac{1}{n}\sum_{k=0}^{n-1} \|S_kC_k\|^2 \tag{8}$$

where $\|S_kC_k\|^2$ denotes the distance between sample S_k and its class center C_k.

Afterwards, the G-loss is composed of two parts, i.e., inter-class loss and intra-class loss:

$$L_G = L_{G-inter} + L_{G-intra} \tag{9}$$

Finally, combining with the traditional cross entropy loss L_c [17], the network is trained with a new loss:

$$L = L_c + \beta L_G \tag{10}$$

where β is a hyper-parameter to control the impact of G-loss.

4. EXPERIMENTS

4.1. Dataset and Settings

The proposed algorithm was evaluated on the H&E stained microscopy images of BACH 2018, i.e., the task of Part A. There are 400 images for training and 100 images for test. Specifically, the images from training set are labelled as normal, benign, in situ carcinoma or invasive carcinoma and each class consists of 100 images with the size of 2048×1536 pixels. In our experiments, the training set was divided by a ratio of 8:2 for training and validation. Note that the labels of the test images are invisible and the test part needs to be done online. Accuracy is considered as the evaluation metric:

$$Accuracy = \frac{\sum_{i=0}^{99} x_i}{100}, \begin{cases} x_i = 1 & (if\ y_i = t_i) \\ x_i = 0 & (if\ y_i \neq t_i) \end{cases} \tag{11}$$

where y_i denotes the predicted label while t_i denotes the ground truth.

In clinical practice, doctors usually use whole-slide images (WSI) for diagnosis. However, the size of the WSI can be as large as $100,000 \times 100,000$ pixels and the microscopy images are parts of WSI, which means it will be time-consuming if our network is too complicated. In addition, the training data are too few and the complex networks have a risk of overfitting. Therefore, we chose light ResNet18

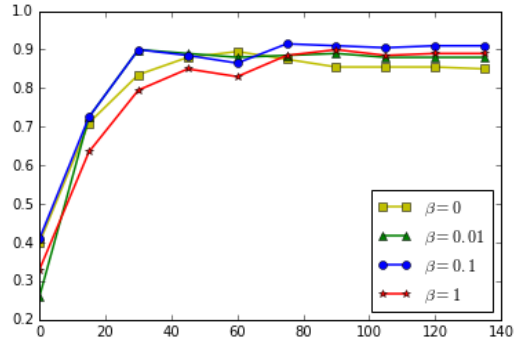


Fig. 7: Accuracy of validation set during training with different β for G-loss.

Table 1: Results of online test

ResNet	HSV	G-inter	G-intra	Accuracy
√				83.7 ± 0.6
√	√			85.0 ± 1.0
√		√		84.3 ± 1.2
√			√	84.3 ± 1.2
√		√	√	86.0 ± 1.0
√	√	√		86.3 ± 0.6
√	√		√	87.0 ± 0.0
√	√	√	√	89.3 ± 1.5

[18] as the skeleton network for fast classification. The training images are first resized to 512×384, then randomly cropped to 384×384, and finally re-resized to 224×224 as the inputs of the network. The basic data augmentation includes flipping and rotating. The learning rate is initialized to 0.001 and will be decreased by a factor of 0.5 after every 20 epochs. The batch size is set as 128. All experiments are conducted based on Pytorch toolbox [19].

The hyper-parameter β for Eq.10 controls the impact of G-loss. Fig.7 shows the accuracy trends on the validation set during training with different β. All experiments implement HSV augmentation. Our network achieves a better result when $\beta = 0.1$.

4.2. Results on Test Set

Table 1 shows online test results and the accuracy (%) is the evaluation metric described in Section 4.1. Due to the randomness of rotation, cropping and scaling during training, each set of experiments is performed three times and the models of epoch 150 are saved for testing. The table presents the mean and standard deviation of every three experiments. In addition to the experiments with or without HSV augmentation (described in Section 3.1), we also conduct experiments on G-loss. The third column of Table 1, i.e., "G-inter" denotes whether G-loss is combined with the inter-class gravitation loss. Similarly, "G-intra" denotes whether G-loss is combined with the intra-class gravitation loss. The hyper-parameter β of Eq.10 is set to 0.1. We can observe that by introducing HSV augmentation and G-loss, the experimental performances are significantly improved.

Table 2: Comparision with other mothods

Rank	Team	Accuracy
1	**Bamboo (Ours)**	**0.91**
2	Sai Saketh Chennamsetty [20]	0.87
2	Scotty Kwow [21]	0.87
4	Nadia Brancati [22]	0.86
5	Matthias Kohl [4]	0.83
6	Hongliu Cao [9]	0.79
7	Gleb Makarchuk [23]	0.76
8	Alexander Rakhlin [24]	0.74
9	Xinpeng Xie [11]	0.72
10	Sulaiman Vesal [3]	0.66
11	Aditya Gokatkar [5]	0.61
11	Kamyar Nazeri [10]	0.61
13	Yeeleng S. Vang [2]	0.47

Table 2 shows the comparison online. All methods are based on the test set of BACH 2018. To be specific, Sai [20] and Scotty [21] shared the champion of the challenge before the deadline. All results submitted before deadline can be obtained [1]. Note that the challenge is re-open for submission and all test results are listed at the website [2]. The leaderboard shows the best performance of every team and our method achieved a very competitive rank.

5. CONCLUSION AND FUTURE WORK

In this paper, we propose a simple but effective classification algorithm for H&E stained breast cancer histology microscopy images. Firstly, to cope with the color variation of stained images, a color augmentation method in HSV color space is applied to increase the color generalization. Then a gravitation loss is proposed to optimize the margins of inter- and intra-class. Compared with the state-of-the-art methods, the proposed algorithm achieves a very promising performance on online BACH 2018 evaluation. Considering that the microscopy images are patches of the whole slide images, in the future work, we will explore a more reasonable solution on the whole slide images.

6. ACKNOWLEDGEMENT

This work is supported by Chinese National Natural Science Foundation (61532018), Key Laboratory of Forensic Marks, Ministry of Public Security of China and BUPT Excellent Ph.D. Students Foundation (CX2019217).

[1] https://iciar2018-challenge.grand-challenge.org/Legacy-results/

[2] https://iciar2018-challenge.grand-challenge.org/evaluation/results/

7. REFERENCES

[1] Freddie Bray, Jacques Ferlay, Isabelle Soerjomataram, et al., "Global cancer statistics 2018: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries," *CA: A Cancer Journal for Clinicians*, 2018.

[2] Yeeleng S. Vang, Zhen Chen, and Xiaohui Xie, "Deep learning framework for multi-class breast cancer histology image classification," pp. 914–922, 2018.

[3] Sulaiman Vesal, Nishant Ravikumar, Amir Abbas Davari, et al., "Classification of breast cancer histology images using transfer learning," in *International Conference Image Analysis and Recognition*, 2018, pp. 812–819.

[4] Matthias Kohl, Christoph Walz, Florian Ludwig, Stefan Braunewell, and Maximilian Baust, "Assessment of breast cancer histology using densely connected convolutional networks," in *International Conference Image Analysis and Recognition*, 2018, pp. 903–913.

[5] Aditya Golatkar, Deepak Anand, and Amit Sethi, "Classification of breast cancer histology using deep learning," in *International Conference Image Analysis and Recognition*, 2018, pp. 837–844.

[6] Ehteshami Bejnordi B, Veta M, Johannes van Diest P, et al., "Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer," *JAMA*, vol. 318, no. 22, pp. 2199–2210, 2017.

[7] Guilherme Aresta, Teresa Araújo, Scotty Kwok, et al., "Bach: Grand challenge on breast cancer histology images," *arXiv preprint arXiv:1808.04277*, 2018.

[8] Babak Ehteshami Bejnordi, Geert Litjens, Nadya Timofeeva, Irene Otte-Hller, et al., "Stain specific standardization of whole-slide histopathological images," *IEEE Transactions on Medical Imaging*, vol. 35, no. 2, pp. 404–415, 2016.

[9] Hongliu Cao, Simon Bernard, Laurent Heutte, and Robert Sabourin, "Improve the performance of transfer learning without fine-tuning using dissimilarity-based multi-view learning for breast cancer histology images," in *International Conference Image Analysis and Recognition*, 2018, pp. 779–787.

[10] Kamyar Nazeri, Azad Aminpour, and Mehran Ebrahimi, "Two-stage convolutional neural network for breast cancer histology image classification," in *International Conference Image Analysis and Recognition*, 2018, pp. 717–726.

[11] Xinpeng Xie, Yuexiang Li, and Linlin Shen, "Active learning for breast cancer identification," *arXiv preprint arXiv:1804.06670*, 2018.

[12] M Tarek Shaban, Christoph Baur, Nassir Navab, and Shadi Albarqouni, "Staingan: Stain style transfer for digital histological images," *arXiv preprint arXiv:1804.01601*, 2018.

[13] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, et al., "Generative adversarial nets," in *Advances in Neural Information Processing Systems*, 2014, pp. 2672–2680.

[14] Tsung-Yi Lin, Priyal Goyal, Ross Girshick, et al., "Focal loss for dense object detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2018.

[15] Fausto Milletari, Nassir Navab, and Seyed-Ahmad Ahmadi, "V-net: Fully convolutional neural networks for volumetric medical image segmentation," in *3D Vision*. IEEE, 2016, pp. 565–571.

[16] I Newton and A.E. Shapiro, "The principia: Mathematical principles of natural philosophy," *Studies in History & Philosophy of Science Part A*, vol. 35, no. 3, pp. 665–667, 2004.

[17] P.T.D. Boer, D.P. Kroese, S Mannor, et al., "A tutorial on the cross-entropy method," *Annals of Operations Research*, vol. 134, no. 1, pp. 19–67, 2005.

[18] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[19] "Pytorch," `http://pytorch.org/blog/`.

[20] Sai Saketh Chennamsetty, Mohammed Safwan, and Varghese Alex, "Classification of breast cancer histology image using ensemble of pre-trained neural networks," pp. 804–811, 2018.

[21] Kwok Scotty, "Multiclass classification of breast cancer in whole-slide images.," in *International Conference Image Analysis and Recognition*, 2018, pp. 931–940.

[22] Nadia Brancati, Maria Frucci, and Daniel Riccio, "Multi-classification of breast cancer histology images by using a fine-tuning strategy," in *International Conference Image Analysis and Recognition*, 2018, pp. 771–778.

[23] Artem Pimkin, Gleb Makarchuk, Vladimir Kondratenko, et al., "Ensembling neural networks for digital pathology images classification and segmentation," in *International Conference Image Analysis and Recognition*, 2018, pp. 877–886.

[24] Alexander Rakhlin, Alexey Shvets, Vladimir Iglovikov, and Alexandr A Kalinin, "Deep convolutional neural networks for breast cancer histology image analysis," in *International Conference Image Analysis and Recognition*, 2018, pp. 737–744.