Deep Multi-Magnification Similarity Learning for Histopathological Image Classification

Songhui Diao, Weiren Luo, Jiaxin Hou, Ricardo Lambo, Hamas A. AL-kuhali, Hanqing Zhao, Yinli Tian, Yaoqin Xie, Nazar Zaki, and Wenjian Qin

***Abstract*—Precise classification of histopathological images is crucial to computer-aided diagnosis in clinical practice. Magnification-based learning networks have attracted considerable attention for their ability to improve performance in histopathological classification. However, the fusion of pyramids of histopathological images at different magnifications is an under-explored area. In this paper, we proposed a novel deep multi-magnification similarity learning (DSML) approach that can be useful for the interpretation of multi-magnification learning framework and easy to visualize feature representation from low-dimension (e.g., cell-level) to high-dimension (e.g., tissue-level), which has overcome the difficulty of understanding cross-magnification information propagation. It uses a similarity cross entropy loss function designation to simultaneously learn the similarity of the information among cross-magnifications. In order to verify the effectiveness of DMSL, experiments with different network backbones and different magnification combinations were designed, and its ability to interpret was also investigated through visualization. Our experiments were performed on two different histopathological datasets: a clinical nasopharyngeal carcinoma and a public breast cancer BCSS2021 dataset. The results show that our method achieved outstanding performance in classification with a higher value of area under curve, accuracy, and F-score than other comparable methods. Moreover, the reasons behind multi-magnification effectiveness were discussed.**

***Index Terms*—Multi-magnification, histopathological image, similarity, classification, deep learning.**

I. Introduction

HISTOPATHOLOGICAL diagnosis is the gold standard for the clinical diagnosis of cancer [1]. Observation of the microscopic lesions structure, cell morphological changes, cancer stage, etc., provides a reference basis for preoperative diagnosis, treatment options, and postoperative prognosis for patients. The development of whole slide images (WSIs) has also brought many applications, such as the recognition of colorectal cancer [2], classification of lung cancer [3], diagnosis of lymphoma [4], and prediction of bladder cancer recurrence [5]. However, pathologists generally rely on experience and knowledge to analyze WSIs [6], which means the evaluation results are subjective. This type of manual analysis is also time-consuming and labor-consuming when dealing with a large quantity of histopathological data [7].

To reduce the burdens of manual analysis, automated analysis techniques have been continuously improved [8], [9], [10]. Utilizing machine learning to analyze WSIs automatically is still a challenge [11]. The development of convolutional neural networks (CNNs) in deep learning has been a powerful new tool. Over the years, numerous methods have been proposed for cancer diagnosis of histopathological images using deep learning and achieved impressive results [12], [13], [14], [15], [16]. Computer-aided diagnosis systems for gastric cancer [13] and lung cancer diagnosis [14] have been successfully implemented on pathological images. For the diagnosis of precancerous lesions of esophageal carcinoma, a semi-automatic classification method based on deep learning has been developed, which can reduce the workload of pathologists by 57% [15]. Although these diagnostic methods have achieved promising results, they all need massive amounts of data, and the images used in the experiment have only a single magnification. In reality, however, pathologists usually combine information at different magnifications [17], i.e., from scales ranging from the sub-nuclear ($\approx$O(0.1 μm)) to the cellular ($\approx$O(10 μm)) and intercellular ($\approx$O(100 μm)) to other higher tissue ($\approx$O(1mm)) sizes, to make diagnoses [18].

In practical application, histopathological images are stored in pyramid form [19], in which each layer has images with different magnifications. In recent years, the research hotspots utilizing the different magnification information in histopathological images mainly focused on multi-scale image processing [20] or

Manuscript received 30 March 2022; revised 14 October 2022 and 7 November 2022; accepted 8 January 2023. Date of publication 16 January 2023; date of current version 7 March 2023. This work was supported in part by the Shenzhen Science and Technology Program of China under Grant JCYJ20200109115420720, in part by the National Natural Science Foundation of China under Grant 62271475, and in part by the Youth Innovation Promotion Association CAS under Grant 2022365. *(Songhui Diao and Weiren Luo contributed equally to this work.) (Corresponding author: Wenjian Qin.)*

Songhui Diao is with the Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences, Shenzhen 518055, China, and also with the Shenzhen College of Advanced Technology, University of Chinese Academy of Science, Shenzhen 518055, China (e-mail: sh.diao@siat.ac.cn).

Weiren Luo is with the Department of Pathology, Shenzhen Third People's Hospital, Shenzhen 518112, China (e-mail: luoweiren@hotmail.com).

Jiaxin Hou, Ricardo Lambo, Hanqing Zhao, Yinli Tian, Yaoqin Xie, and Wenjian Qin are with the Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences, Shenzhen 518055, China (e-mail: jx.hou@siat.ac.cn; ricardo@siat.ac.cn; hq.zhao@siat.ac.cn; liying2006007@163.com; yq.xie@siat.ac.cn; wj.qin@siat.ac.cn).

Hamas A. AL-kuhali is with the School of Computer and Artificial Intelligence, Wuhan University of Technology, Wuhan 430070, China (e-mail: kuhamas0@gmail.com).

Nazar Zaki is with the College of Information Technology, United Arab Emirates University, Al Ain 15551, UAE (e-mail: nzaki@uaeu.ac.ae).

Digital Object Identifier 10.1109/JBHI.2023.3237137

multi-scale feature utilization [21]. This paper distinguishes the difference between multi-scale and multi-magnification images in the A part of the Method. In multi-scale image processing, Nishio et al. [22] extracted the features of different scale images by statistical methods and classified them by machine learning. As for deep learning methods, Chen et al. [23] added a multi-scale framework to the vision transformer network. The primary way to obtain multi-scale images is to apply different cropping methods to the same image to obtain images with different resolutions or different visual fields. Tong et al. [24], [25] adopted a similar strategy in the classification of breast cancer. Marini et al. [26] proposed a multi-scale-task learning convolutional neural network to diagnose colon cancer in which the size and resolution of images used in WSI combinations are the same, though their scales are different. Each image is a subsection of a different image, magnified until they are the same size. Sun et al. [27] proposed a method to learn differently scaled images and classify histopathological images by simultaneously inputting them into the network. The image pairs input into the network come from different magnifications, and the information of the image pairs is different. Among the methods using the multi-scale features, there are many exciting works. Gao et al. [28] proposed a new CNN block representing multi-scale features at a granular level, which means the increase in perception is quite a small level for each network layer. Zhou et al. [29] jointly modulated features using features at different scales in high-dimension feature space, then gradually unsampled and refined them until their final desired resolution. In the present work, only image information at one magnification was input into the network, while multi-scale features are incorporated in the use of the image characteristics at different scales. Making full use of multi-scale features is equivalent to adding some learnable parameters modules to the original network, which is essentially different from the direct multi-magnification image.

The abovementioned techniques using multi-scale images and multi-scale features for processing differ from multi-magnification methods in which the network processes different magnifications of the same image. Since sufficient information for multi-magnification classification is available, a multi-magnification learning strategy inspired by the approach of pathologists is proposed in the current study. The multi-magnification work already in the literature that uses deep learning methods can be roughly divided into two research directions. In the research field of machine learning, Doyle et al. [30] proposed to input multi-magnification images to a Bayesian classifier for classification to achieve prostate cancer detection in 2010, but its classification performance was not outstanding. As the first example of the deep learning method, Chen et al. [31] proposed a network feeding multi-magnification images to the shared depth network for semantic segmentation. The attention mechanism was used to weigh the characteristics of different magnification image features. Das et al. [32] also proposed a deep CNN framework, which analyses images from a random number of regions of the tissue section at multi-magnification. Diagnosis at the slide level is then processed using a majority voting-based approach. These researches show that multi-magnification information is also effective for diagnosing the histopathological image, which is more in line with doctors' actual diagnostic procedures. However, these studies are trying to eliminate the effects of different magnification, which was equivalent to a data augmentation strategy. Research in the other direction mainly focuses on feature fusion [33], [34], [35]. Such as the work of Tao et al. [34], which used an attention-based method to combine the features of multi-magnification images for prediction. Lin et al. [35] have demonstrated a multi-magnification architecture called Feature Pyramid Network for constructing high-level semantic feature maps at all scales. In our previous work [36], a multi-magnification histopathological image with weakly supervised attention detection framework based on CNN was proposed, in which images with different magnifications were given different weights. However, these studies mainly focus on how to utilize multi-magnification image features. This may cause different magnification images to affect each other when extracting features and even cause learning in a direction such that the optimization goal is more challenging to achieve. Moreover, there is no reasonable explanation for why multi-magnification methods perform better than multi-scale image and multi-scale features. Inspired by the work of Sun et al. [27], we further analyze the above situation by evaluating the similarity of the features of images at different magnification.

Similarity evaluation is a standard image evaluation method [37], [38], which can also be applied to medical images [39], [40], [41]. This framework extracted structural information from a scene based on the degradation of structural information to develop a Structural Similarity Index. We first learn local and global information through a multi-magnification backbone network, in which the highest-level feature vector output is to learn the same classification target in a direct way rather than complex strategies and parameters. The classification results are obtained by fusing the prediction information at different magnifications. Since the information functions are different at various magnifications, similarity is used to evaluate the availability of information. Our method is validated by quantitative results in Section III and by quantitative and qualitative analysis in Section IV.

The main contributions of this paper are as follows:

1) We proposed a novel similarity learning approach that can be useful for the interpretation of multi-magnification learning framework and easy to visualize feature representation from low-dimension (e.g., cell-level) to high-dimension (e.g., tissue-level), which has overcome the difficulty of understanding cross-magnification information propagation. The similarities of low-dimension and high-dimensional information at different magnifications are characterized by similarity theory.
2) The effectiveness of multi-magnification information is further explained by analyzing the similarity. The statistics of different image pair classification results, especially their misclassification, can give direct evidence of how well the use of multi-magnification information works. Furthermore, the region of interest (ROI) is found and tracked using multi-magnification images to explain the working mechanism of multi-magnification information learning.
3) The proposed method achieves superior performance compared with state-of-the-art methods on clinical and public datasets.

II. Method

This paper proposes a novel framework for histopathological image analysis, which is useful for histological diagnosis. This section describes the definition and characteristics of

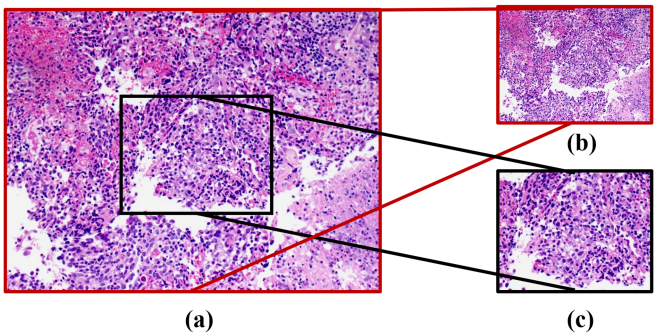

Fig. 1. Examples of each type of image pair. (a) and (b) refer to multi-magnification image combinations (marked in red). (a) and (c) refer to multi-scale image combinations (marked in black).

multi-magnification histopathological images. We then expound on the proposed framework (DMSL), an improved classification network, with Inception_v3 [42] as the backbone, and also describe the two main parts of the framework: the training and implementation schema. Briefly described, that framework is a deep CNN based on multi-magnification images, including a similarity evaluation module to constrain the optimal features under different magnification images. In order to learn information at different magnifications effectively, a specific loss function is designed. Furthermore, we investigate the behavior and efficiency of the proposed framework by visualizing what it has learned from multi-magnification images.

A. Multi-Magnification Histopathological Image

Histopathological images are stored with different magnifications in a WSI, and these images contain different information and resolution. Our method is based on the multi-magnification histopathological image, which is different from multi-scale images and multi-scale features. In this paper, multi-magnification images refer to images with different resolutions in the same field of view (FOV), while multi-scale images have different FOVs at the same or different resolutions. The method of multi-scale features [43] is thus different from those of the two we mentioned earlier. An example of each type of image pair is displayed in Fig. 1.

After discussion with pathologists from Shenzhen Third People's Hospital and investigating related papers [44], [45], We can see that the FOV of the pathologist inspecting at the WSI under the microscope at different resolutions is relatively fixed. Even if the FOV of the pathologist changes at different magnifications, the pathologist usually observes the same area near the center of the field.

B. Proposed Multi-Magnification Framework

1) Basic Network Architecture: Our basic network is designed based on the Inception_v3 structure, which is widely used in medical image classification tasks and shows lots of advantages. It adopts asymmetric decomposition convolution and convolution substitution, which reduces the amount of calculation and increases the diversity of features. Its Inception Reduction Module block strengthens the adequate transmission of information by expanding the filter banks, giving rise to more accurate classification results.

An overview of the framework is given in Fig. 2. We have removed the last convolution layer and all subsequent fully connected layers. Therefore, the remaining structure constitutes the encoder for feature extraction, which was learnable rather than fixed parameters.

Two images with different magnifications, magnification 1 and magnification 2, are fed into two separate backbones, respectively. Then, the features obtained are input into two branches. The first branch is a classifier consisting of fully connected layers and SoftMax layers to obtain the classification probability of the current magnification image. Finally, for every tuple of images, the results of the fusion classifier are obtained according to (1):

$$Output = \sum_{m} \omega m classifier m(fm) \tag{1}$$

where fm represents the feature of the magnification m, ωm represents the weight of the classification result of m, and the sum is one. Another branch is a similarity measurement module (SIM) which evaluates the extracted features from different magnification images. The similarity is computed based on the L1-Norm of the two or more features maximized value on the channel of feature:

$$SIM = \sum_{2\le i\le m, 1\le j\le i} \left| \sum_{k=1}^{h*w} \max_{C} X_{i_k}^{C} - \max_{C} X_{j_k}^{C} \right| \tag{2}$$

where $X_{i_k}^{C}, X_{j_k}^{C}$ represents the feature in location k of channel C under the magnification m, and where h and w represent the height and width of the image feature map. In order to calculate the different magnification features with different shapes, we changed different features into the target same shape by adaptive pooling [46].

2) Loss Function: Cross entropy (CE) is a loss function commonly used in classification tasks. In our training scheme, for searching the tissue of common concern in different magnifications, the SIM is added to the loss function:

$$Loss_{ALL} = \alpha Loss_{CE} + \beta Loss_{SIM} \tag{3}$$

where $Loss_{CE}$is the CE of the predicted label and ground truth (GT), $Loss_{SIM}$ is the absolute value of SIM, hyper-parameters, $\alpha$and β , are the weight for each loss and are set by experience. The optimization objective of the model is the minimum value of $Loss_{ALL}$.

C. Network Training and Implement

1) Data Strategy: In the training schema, the diagnostic model is trained by normalizing and enhancing histopathological images of each tuple. Usually, observation of various angles will happen in the pathologists' diagnosis. We adopted many augmentation methods to fully consider the visual features that pathologists may encounter under the microscope in the augmentation processing. Meanwhile, we simulate pathologists' examination processes and authenticity in the actual diagnosis, such as color jitter, random rotation, and gray-scale transformation.

In general, the category of the histopathological image containing cancer is that of the tumor. Considering the actual training and learning of efficient information, images with a low proportion of cancer would not be selected. More than 80% of the images with cancer regions are tumor data, and the images

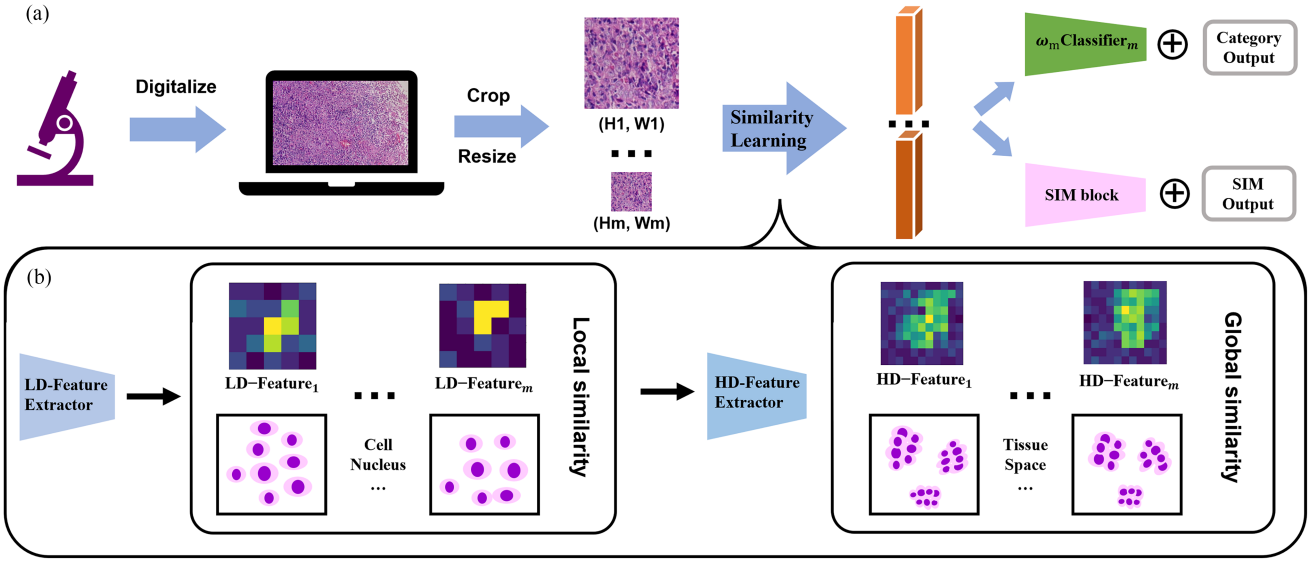

Fig. 2. Overview of the proposed DMSL. The part (a) shows our proposed block diagram. WSIs were obtained by scanning and digitizing tissue samples. Then the different magnification images obtained by cropping and resizing were fed into the similarity learning network (part b). The obtained feature vector (orange mark) entered the following two branches. One branch flattens the features of different magnification images, and then aggregates the prediction results through a weighted classifier ($\omega mClassifierm$). Another branch constrains the similarity of the features of different magnification images (SIM block). The part (b) consisted of low dimensional feature extraction (LD-Feature Extractor$_{\mathrm{m}}$, the *m* means magnification) and high-dimensional feature extraction (HD-Feature Extractor$_{\mathrm{m}}$), in which the structure of LD and HD-Feature Extractor are different parts of the backbone network, inception_v3. The part from input to the first convolution layer was LD-Feature Extractor to learn local similarity such as morphological feature of cell and nucleus, and the part from the leftover convolution layer to the classifier is HD-Feature Extractor to learn global similarity such as tissue distribution and spatial pattern. The preliminary similar features obtained from LD-Feature Extractor enter HD-Feature Extractor to compute the similar features with global scope.

with no cancer regions are normal data. The above is the dataset selection strategy for different categories.

2) Transfer Learning: One of the most significant benefits of transfer learning is that researchers can skillfully apply the knowledge learned before to solve new problems better or faster [47]. The properties of transferring knowledge between different tasks and data domains can effectively reduce the burden for model retraining. In medical image analysis based on deep learning, it is usually necessary to have enough annotated data. In practice, medical image annotation is costly and time-consuming work. Transfer learning can be used in medical image analysis to reduce the negative impact of insufficient data. Many studies [47], [48], [49] have shown that the application of transfer learning in medical image analysis can improve performance to a certain extent, even if some medical images are different from natural images in texture and structure. Consequently, transfer learning has also been added to our training strategy.

D. Visualization of DMSL

A visualization procedure is executed to intuitively verify whether the proposed DMSL effectively makes it interpretability. The intermediate results of multi-magnification related network modules are visualized to help us better understand their working mechanism. In order to validate whether the depth network can learn similar essential information in different magnification images, we generate heat maps through the gradient information in the convolution layer of the low dimension (the first layer) and high dimension (the last layer) feature stage, which was inspired by the Grad-CAM [50]. Analogous to the thermal image generated by thermal imaging equipment using the infrared principle [51], [52], Grad-CAM firstly obtains the gradient of the feature map according to the backward direction of the output vector and obtains the gradient corresponding to each pixel on each feature map, that is, the gradient map corresponding to the feature map. The mean of each gradient graph is then calculated, which corresponds to the weight of each feature graph. Finally, the final class activation graph can be obtained through the linear rectification activation function [53] after the weights are used in a weighted sum with the feature graph.

III. Experiments and Results

This section will comprehensively evaluate the classification performance of the proposed DMSL in the histopathological image. The following multi-magnification image experiments were based on two feature extractors and two classifiers, that is, there were two networks rather than one common network.

A. Datasets

The following two datasets are used to validate the DMSL model, of which one is a public dataset, and the other is a clinical dataset. The reason for choosing these datasets is that there is a maximum magnification of 40x, which allows us to carry out various experimental combinations of multi-magnification. In addition, the combination of public and clinical datasets can better verify our model performance. The experiment is mainly performed on the clinical dataset and further tested on the public dataset as an external dataset.

NPC2020 [54]: This clinical dataset consists of 608 patients, which are collected in the pathology department of Gaozhou people's Hospital and Shenzhen Third People's hospital.

TABLE I
DETAIL OF THE NPC2020 DATASET WITH VARIOUS MAGNIFICATION

Classes	Training			Testing			Total		
	Tumor	Lymphoid hyperplasia	Inflammation	Tumor	Lymphoid hyperplasia	Inflammation	Tumor	Lymphoid hyperplasia	Inflammation
Cases	98	83	231	74	41	81	172	124	312
Slides	470	130	644	312	85	421	782	214	1065

Inflammation refers to chronic nasopharyngeal inflammation.

TABLE II
DETAIL OF THE BCSS2021 DATASET WITH VARIOUS MAGNIFICATION

Classes	Training	Testing	Total
Tumor	1.47×10^4 (76)	0.44×10^4 (24)	1.92×10^4 (100)
Non-tumor	1.63×10^4 (74)	0.58×10^4 (26)	2.21×10^4 (100)

Data are expressed as n (p), while n means the number of patches and p means the percentage of the total patch number.

TABLE III-A
COMPARISON OF OUR MODELS TO EXISTING METHODS ON THE NPC2020 DATASET

Method	Combination	Mode	NPC2020 dataset (mean ± std)				
			PRE	REC	ACC	F1	AUC
Vgg19	20x	/	0.854±0.024	0.801±0.014	0.865±0.014	0.827±0.014	0.903±0.025
	40x	/	0.850±0.032	0.796±0.018	0.854±0.028	0.822±0.015	0.885±0.016
Resnet50	20x	/	0.924±0.005	0.820±0.014	0.909±0.009	0.859±0.011	0.942±0.008
	40x	/	0.902±0.013	0.821±0.013	0.903±0.012	0.852±0.005	0.937±0.008
Inception_v3	20x	/	0.843±0.044	0.818±0.011	0.871±0.023	0.827±0.026	0.936±0.006
	40x	/	0.916±0.007	0.812±0.013	0.902±0.006	0.851±0.010	0.937±0.004
Shufflenet_v2	20x	/	0.877±0.012	0.804±0.009	0.891±0.007	0.834±0.010	0.929±0.007
	40x	/	0.914±0.009	0.827±0.008	0.911±0.005	0.861±0.006	0.943±0.005
Mnasnet	20x	/	0.896±0.067	0.816±0.037	0.907±0.205	0.853±0.019	0.936±0.012
	40x	/	0.894±0.069	0.826±0.104	0.906±0.203	0.862±0.072	0.944±0.022
Regnet	20x	/	0.902±0.009	0.801±0.019	0.892±0.010	0.838±0.016	0.936±0.010
	40x	/	0.893±0.009	0.797±0.012	0.890±0.006	0.833±0.009	0.943±0.008
MSCN	40x+20x	MS	0.898±0.012	0.847±0.017	0.904±0.014	0.865±0.016	0.963±0.004
Ours	40x+20x	MM	**0.956±0.004**	0.870±0.015	0.933±0.008	0.904±0.010	0.980±0.003
	40x+20x	SIM	0.943±0.009	**0.882±0.012**	**0.937±0.007**	**0.909±0.010**	**0.984±0.004**

The best results for each metric are shown in bold, and the second are underlined.

The collection and use of these datasets obtained informed consent from the Institutional Research Ethics Committee. All slides collected are divided into three categories: inflammation (Imf), lymphoid hyperplasia (Lym), and nasopharyngeal carcinoma (NPC, Tum) diagnosed as nonkeratinizing carcinoma according to the histological classification of the World Health Organization. Specifically, these slides are jointly labelled by two pathologists with at least fifteen years of experience, and possible conflicting annotations have been negotiated without consensus. The image patches are separated into training datasets and testing datasets in a suitable proportion. The details of dataset division are shown in Table I.

BCSS2021 [55]: This dataset (publicly available at https://github.com/PathologyDataScience/BCSS) contains 151 hematoxylin and eosin-stained WSIs with breast cancer for semantic segmentation. Together with the original WSI, it comes from The Cancer Genome Atlas. The annotation information is taken from the literature. Specifically, a study coordinator, a medical doctor, selects one representative ROI with a mean size of 1.18mm^2, labelling different regional boundaries of tissue and its category in each slide, such as a tumor, adipose, blood vessels, etc. It should be noted that not every slide has all-category labels, but tumor slides do. All images are divided into only two categories, tumor (Tum) and non-tumor (Oth), considering the expandability and complexity of the experiment. All slides were used in our experiment, and the details of dataset division are shown in Table II. The image patches are also based on the principle of independence.

B. Implementation Details

A simple general method is used to normalize the image through mean and variance. As for the way to determine the patch location in each case, the selected strategy is random, and the random number is set for tiles. For patches from tiles, the patches were cropped with 0.25 overlap of patch size. Note that if the proportion of tumor area is not less than 80%, the category of the patch is the tumor, while the proportion of tumor area must be 0% to be judged as non-tumor when judging the category in the previous step, according to the annotations by pathologists. Moreover, the patch resolution of 40× is 600 pixels by 600 pixels, and the 20× image resolution is 300 pixels by 300 pixels.

TABLE III-B
COMPARISON OF OUR MODELS TO EXISTING METHODS ON THE BCSS2021 DATASET

Method	Combination	Mode	BCSS2021 dataset (mean ± std)				
			PRE	REC	ACC	F1	AUC
Vgg19	20x	/	0.833±0.010	0.884±0.010	0.869±0.010	0.849±0.010	0.936±0.006
	40x	/	0.826±0.069	0.885±0.103	0.867±0.016	0.850±0.017	0.935±0.012
Resnet50	20x	/	0.853±0.019	0.878±0.021	0.886±0.016	0.863±0.019	0.944±0.010
	40x	/	0.853±0.011	0.873±0.015	0.886±0.009	0.861±0.011	0.941±0.010
Inception_v3	20x	/	0.863±0.015	0.881±0.009	0.894±0.011	0.871±0.012	0.945±0.004
	40x	/	0.858±0.007	0.873±0.008	0.890±0.004	0.865±0.004	0.944±0.006
Shufflenet_v2	20x	/	0.814±0.036	0.850±0.022	0.844±0.035	0.821±0.032	0.926±0.018
	40x	/	0.836±0.008	0.853±0.002	0.870±0.005	0.842±0.009	0.939±0.005
Mnasnet	20x	/	0.816±0.011	0.845±0.017	0.854±0.021	0.829±0.009	0.901±0.018
	40x	/	0.830±0.010	0.852±0.004	0.871±0.016	0.841±0.008	0.915±0.047
Regnet	20x	/	0.812±0.015	0.876±0.015	0.840±0.021	0.823±0.021	0.933±0.007
	40x	/	0.844±0.019	0.879±0.010	0.879±0.017	0.857±0.019	0.940±0.005
MSCN	40x+20x	MS	0.866±0.008	0.888±0.012	0.897±0.007	0.875±0.009	0.945±0.002
Ours	40x+20x	MM	0.877±0.004	0.903±0.008	0.907±0.003	0.888±0.004	0.956±0.003
	40x+20x	SIM	**0.878±0.006**	**0.905±0.008**	**0.909±0.005**	**0.890±0.007**	**0.959±0.003**

These tables records the results of training the model using the four different strategies seen in the Mode column: MS means the model is trained with multi-scale images, MM means the model is trained with multi-magnification images, '/' means model is trained based on single magnification, and SIM means the model is trained with multi-magnification images with SIM.
The best results for each metric are shown in bold, and the second are underlined.

TABLE IV
COMPARISON OF EXISTING METHODS USING OUR FRAMEWORK IN TWO DATASETS, WHICH THE MAGNIFICATION COMBINATION USED WAS 20×+40×

Experiments	Mode	NPC dataset			Breast dataset		
		ACC	F1	AUC	ACC	F1	AUC
Vgg19	MM	0.891±0.004	0.865±0.008	0.959±0.005	0.884±0.004	0.865±0.006	0.917±0.005
Resnet50	MM	0.921±0.003	0.886±0.004	0.964±0.007	0.896±0.003	0.877±0.004	0.946±0.004
Ours	MM	0.933±0.008	0.904±0.010	0.980±0.003	0.907±0.003	0.888±0.004	0.956±0.003
Vgg19	SIM	0.894±0.004	0.867±0.011	0.963±0.004	0.889±0.005	0.867±0.005	0.919±0.006
Resnet50	SIM	0.923±0.005	0.888±0.008	0.971±0.003	0.901±0.002	0.883±0.002	0.955±0.003
Ours	SIM	**0.937±0.007**	**0.909±0.010**	**0.984±0.004**	**0.909±0.005**	**0.890±0.007**	**0.959±0.003**

The best results for each metric are shown in bold.

Besides, the hyperparameters of weight in (1) were set to 0.5, respectively, and so are α andβ.

All models were implemented using PyTorch (version 1.9.0), and all training processes were trained on the NVIDIA RTX A6000 GPU in Linux (version 4.4.0–116-generic). In addition, all experiments had 5-fold cross-validation. The initial learning rate, batch size and optimizer are 0.001, 32, and Adam, respectively.

Precision (PRE), recall (REC), accuracy (ACC), F-score (F1), and area under curve (AUC) were used to evaluate the performance of different models. It should note that the following statistical metrics are based on patch level because the value of metrics based on patient-level is almost 1 with no necessity of comparison, especially in the nasopharyngeal carcinoma dataset.

C. Classification Performance and Comparison Results

In this section, several experiments are performed to prove the effectiveness of the proposed framework.

Tables III and IV show the quantitative classifications patch-wise results of our proposed method and other state-of-the-art methods for 20x and 40x magnifications factors. In this section, our results are based on different combinations of 20x and 40x, including multi-magnifications combined and multi-magnifications combined with SIM bases on two backbones. Among the state-of-the-art deep learning models for image classification, Vgg19 [56], Resnet50 [57], Inception_v3 [42], Shufflenet_v2 [58], Mnasnet [59] and Regnet [60], which obtain good experimental results in classification in recent years, are compared with our DMSL method. Meanwhile, whether it is close to the depth of the proposed model is the main factor used to select the number of layers of the network. Furthermore, according to [61], we reproduced a work on a multi-scale image classification network called MSCN in our experiment. This work also uses double branches to extract image features with different magnification, but the resolution of different magnification images is different. The backbone if MSCN is Vgg16 or Inception_v4 in the original. For the fair comparison, we use the backbone network used by the proposed framework to replace the backbone network of MSCN. In the following experimental results, the experiment on the dataset of NPC2020 is a three-classification task, while the experiment on the dataset of BCSS2021 is a two-classification task. Therefore, the macro

method is adopted for the statistical metrics of the first dataset, which means calculating metrics for each label and finding its unweighted mean without consideration of label imbalance.

In the testing dataset of NPC2020, the performance of the proposed DMSL framework outperforms the classical single magnification models. The improvement rate of AUC goes up by 3.6% to 9.5% in combinations of 20x and 40x. Specifically, from Table III, the performances of the DMSL model with SIM are the best, with the highest evaluation on AUC (0.984), ACC (0.937), REC (0.882), and F1 (0.909). Compared with the above state-of-the-art network results, DMSL without SIM increases by 3.8% to 7.7% on AUC and by 4.5% to 7.7% on F1 at 20x magnification, and by 3.6% to 9.5% on AUC and by 4.2% to 8.2% on F1 at 40x magnification. In the single magnification classification results, the values of each metric of each model under 40x are higher than those under 20x, except for Vgg19 and Resnet50, and the values of the state-of-the-art network generally have low REC values. Our proposed DMSL increases the REC by 9% on average, which greatly improves the recognition accuracy of the tumor category. Overall, our method using multi-magnification has achieved promising results.

For the BCSS2021 dataset, the classification performance of all models is encapsulated in Table IV in terms of the testing set. The proposed DMSL also outperforms classical single magnification model results in which the improvement rate of each metric is from 1% to 5.8% on average in combinations of 20x and 40x, which indicates the effectiveness of the proposed method evaluated using an external dataset. Specifically, from Table III, the performances of the DMSL model with SIM are best, giving the highest values of AUC (0.959), ACC (0.909), PRE (0.878), REC (0.905), and F1 (0.890). DMSL without SIM increases by 1.1% to 5.5% in AUC and 1.7% to 6.7% on F1 at 20x magnification, and 1.2% to 4.1% in AUC and 2.3% to 4.7% in F1 at 40x magnification. While the recognition rate of the tumor category is improved, the recognition rate of the non-tumor category is also greatly improved. Experimental results show that our method has good robustness, combining the performance of the previous dataset. Meanwhile, the dataset volume of this dataset is 5% of that of the previous dataset, which shows that the proposed framework can also achieve better performance on a small volume of data.

Compared with MSCN, our method improves AUC by 1.4%-2.1% in both datasets. The original results of all the above methods are shown in Table III and Table IV. Through further comparison, it can be found that the effectiveness of multi-magnification histopathological image fusion is better in both single backbone and double backbone in different datasets. The conclusion can be drawn that the combination of 20x and 40x magnification is instrumental in discriminating tumors. Moreover, it can be seen that there is little difference between the performance of DMSL and DMSL without SIM, i.e., 0.4%-0.7% on AUC.

D. Results of Different Backbones

In order to verify the adaptability of our framework, we use different backbone networks for feature extraction. More representative Vgg19 and Resnet50 are added as backbone networks for experiments. Han et al. [62] achieved the best results in thyroid pathological image recognition combined with 20x and 40x. Our experiments also use this combination, including the experiment in the last section. The results using different network backbones are shown in Table V. It can be seen that the classification results using the similarity constraint are improved to varying degrees. Moreover, our framework makes the standard deviation smaller, and the performance of that is more stable. As shown in Table V, the result of using Inception_v3 as the network backbone is the best. Compared with the other two feature extraction backbone networks, Inception_v3 has fewer parameters, a faster learning speed, and requires less memory. Therefore, Inception_v3 is the backbone network of all experiments.

E. Results of Different Combinations

To evaluate the effectiveness of the proposed method, different combinations of magnification and different combinations of the method are carried out on two datasets. As shown in Table IV, six combinations of different magnifications are applied to the experiment. They are: $40\times + 20\times$, $40\times + 10\times$, $40\times + 5\times$, $20\times + 10\times$, $20\times + 5\times$ and $10\times + 5\times$. The mean and standard deviation of different magnification combinations intuitively show that DMSL achieved stable performance in different magnification combinations. No matter which combination is adopted, adding an extra magnification of image information can significantly improve the single magnification classification results. It can be seen that the combination of $5\times + 20\times$ achieved better results in both datasets. In addition, the significance of adding similarity constraints will be discussed in the next section, with the combination of $5\times + 20\times$ as an object of further research.

IV. Discussion

In this study, we proposed a novel framework (DMSL) for classifying histopathological images using a new deep convolution model based on multi-magnification. Several experiments were carried out on two different datasets to verify our method, and we obtained encouraging results. However, the reason for the effectiveness of the multi-magnification method has not been discussed in previous work. The following discussion mainly focuses on the DMSL with multi-magnification images.

The direct function of the multi-magnification image provides more information, which is conducive to the network learning more features at each iteration. Generally, high-dimensional features often represent the essential information of the image in the classification task. Evaluating the correlation of high-level features thus becomes feasible. Next, we will analyze the reasons from several angles based on their similarity.

A. Dataset Volume

In our proposed network structure, a double branch network extracts image features at different magnifications on each branch, which means that the input information is doubled for the whole network framework. A question worth asking is whether our model performance can reach that of the single magnification even if the proportion of the training dataset is reduced to half or less. For that, we experimented with the training data set screening to verify our idea using two datasets: the experimental results of the 5x and 20x combination with SIM are shown in Fig. 3.

When the proportion of the training dataset is reduced to 50%, a very competitive result is still achieved, i.e., the AUC is only reduced by 0.7% in NPC2020 and by 0. 4% in BCSS2021.

TABLE V
RESULTS OF DIFFERENT COMBINATION MAGNIFICATION AND WHETHER SIMILARITY CONSTRAINTS WERE USED IN TWO DATASETS

Experiments	Mode	NPC dataset			Breast dataset		
		ACC	F1	AUC	ACC	F1	AUC
20x+40x	MM	0.933±0.008	0.904±0.010	0.980±0.003	0.907±0.003	0.888±0.004	0.956±0.003
10x+40x	MM	0.927±0.004	0.895±0.007	0.973±0.005	0.906±0.003	0.887±0.003	**0.956±0.003**
5x+40x	MM	0.930±0.008	0.899±0.013	0.971±0.007	0.908±0.008	0.889±0.009	0.955±0.006
10x+20x	MM	0.932±0.006	0.909±0.010	**0.985±0.006**	0.908±0.003	0.890±0.003	0.955±0.005
5x+20x	MM	**0.936±0.006**	**0.911±0.008**	0.981±0.008	**0.909±0.005**	**0.892±0.005**	0.954±0.003
5x+10x	MM	0.923±0.006	0.891±0.007	**0.981±0.004**	0.889±0.004	0.867±0.004	0.931±0.007
20x+40x	SIM	0.937±0.007	0.909±0.010	0.984±0.004	0.909±0.005	0.890±0.007	0.959±0.003
10x+40x	SIM	0.925±0.005	0.891±0.009	0.970±0.008	0.908±0.004	0.890±0.004	0.960±0.003
5x+40x	SIM	0.932±0.006	0.902±0.007	0.977±0.002	0.912±0.003	**0.899±0.003**	**0.963±0.003**
10x+20x	SIM	**0.940±0.005**	**0.916±0.009**	0.986±0.003	0.912±0.005	0.895±0.005	0.960±0.003
5x+20x	SIM	0.937±0.005	0.914±0.005	**0.987±0.004**	**0.913±0.003**	0.898±0.003	0.961±0.002
5x+10x	SIM	0.930±005	0.902±0.006	0.963±0.007	0.891±0.009	0.870±0.009	0.929±0.004

The best results for each metric are shown in bold, and the second are underlined.

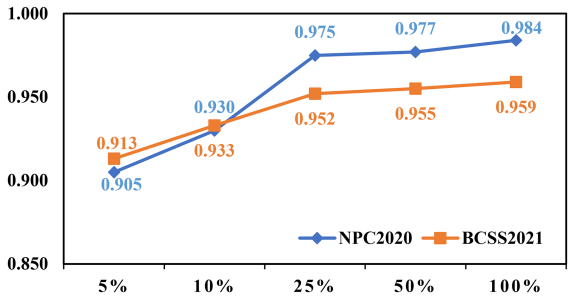

Fig. 3. AUC values for datasets of different proportions in two datasets. The x-axis represents the training dataset ratio, and the y-axis represents AUC value.

However, this result may be due to the use of a double branch network structure by which DMSL has nearly double the number of parameters to learn the feature of the image. The above experiments cannot directly prove that multi-magnification images cause it. Therefore, experiments with fewer data are also carried out. Our model achieves competitive results when the proportion of training datasets is reduced to 10% compared with some advanced models based on single magnification. As shown in Fig. 3, impressive results are also obtained when the proportion of training datasets is reduced to only 5%.

The preliminary conclusion can be drawn that the multi-magnification image can effectively improve the learning ability of the network and the performance of classification.

B. Similarity Statistics

The pathologist zooms in at high magnification after finding the ROI at low magnification, for which the field of vision is the same under the microscope. This is the main factor in the effectiveness of using multi-magnification based on similarity. From the current experimental results, one can see that the multi-magnification image improves the performance of the classification model effectively. However, it cannot be ignored that the improvement of the model with similarity constraints is minor. We suspect the reason is that the multi-magnification model, without increasing the similarity constraint, is actually learning the similarity of different magnification images.

To prove our assumption, an experiment is designed to separately count the similarity of each pair of multi-magnification images when classifying whether it contains similarity constraints in two datasets. Specifically, for each group of experiments, we divide the similarity into different ranges: larger dissimilarity, smaller dissimilarity, smaller similarity, and larger similarity. The corresponding similarity pairs were 0–0.25, 0.26–0.5, 0.51–0.75 and 0.76–1. The value of the similarity was calculated by

$$S(x,y) = \frac{1}{2} \times \left(\frac{(x-\bar{x})^T (y-\bar{y})}{\|x-\bar{x}\| \cdot \|y-\bar{y}\|} + 1 \right), \tag{4}$$

where $\|x\| = (x^T x)^{1/2}$, $\|y\| = (y^T y)^{1/2}$, $\bar{x}$ and $\bar{y}$ are the mean of x and y according to the normalized correlation [63]. The first step is to calculate the similarity for each pair of multiple magnified images in the test dataset. The second step is to count the number of images in each similarity interval. It should be added that in order to ensure the consistency of statistical standards, the statistical objects were the images correctly classified by the two models, i.e., TP and TN, no matter what category. As shown in Fig. 4, larger similarity and smaller similarity images account for more than 90% in the two datasets, respectively, whether similarity constraints are used or not.

Moreover, the whole similarity distribution with similarity constraints and the whole distribution without similarity constraints have a relatively high overlap. Compared with the model without similarity constraints, the model with similarity constraints can improve the similarity of some low-similarity multi-magnification image pairs. The results show that the larger dissimilarity ratio of multi-magnification was no bigger than 3%, whether similarity constraints were used or not in different datasets.

The above statistical experimental results are based on correct classification. However, for some images that are not classified correctly, i.e., FP and FN, the performance of the similarity of these images also needs to be studied, and relevant experiments have also been carried out. Fig. 4 shows that the proportion of larger dissimilarity images and smaller dissimilarity images is close to 80% in the two datasets, respectively, whether similarity constraints are used or not. In addition, the overlap between the whole distribution with similarity constraints and the whole

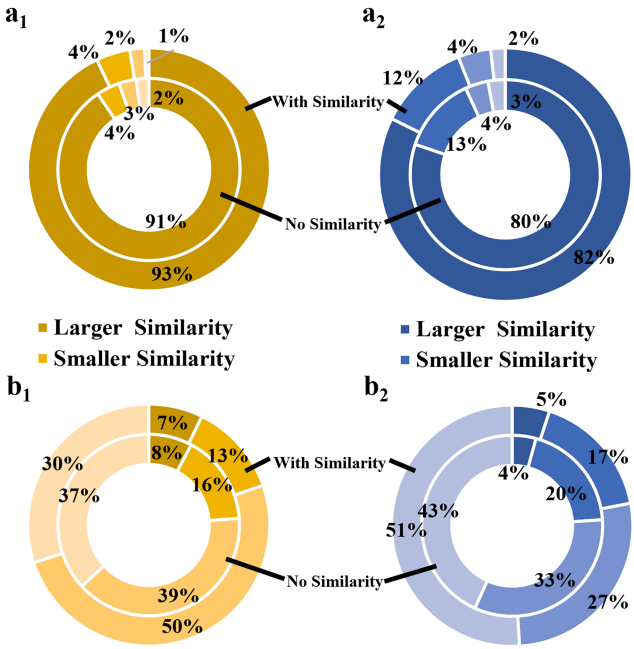

Fig. 4. Similarity statistics for each pair of multi-magnification images, where row (a) represents the statistics of correct classification and row (b) represents the statistics of the wrong classification, subscript 1 represents the NPC2020 dataset, and subscript 2 represents the BCSS2021 dataset, and the inner pie chart represents a model without similarity constraints, and outer pie chart represents a model with similarity constraints.

distribution without similarity constraints is close. It can be further concluded that the model mainly learns similar information between the two images when learning the information of multi-magnification images.

C. Visualization

To explain the reason for improvement and how similarity works in actual network learning, we carried out experiments on the visualization research of model learning. Precisely, two convolution layers in the network, i.e., the convolution layer of low dimension (the first layer) and high dimension (the last layer), were calculated for Grad-CAM at different magnifications in the learning process. Fig. 5 gives examples of heat maps and their corresponding original image patches generated for three categories of the NPC2020 dataset. Fig. 6 gives examples of heat maps and their corresponding original image patches generated for two categories of the BCSS2021 dataset. For the convenience of the display, the images were scaled to the same size, although the image resolution of each magnification was different. It can be seen that the proposed DMSL was able to apply information related to different magnifications. On the heat map at the low to the high stage of DMSL, the prominent areas are gradually clustered, and the proportion of heat map areas with high magnification is higher, indicating that the areas considered by different magnifications are similar, and more detailed features can be learned from high magnification images. These visualization results intuitively show that the proposed network architecture can use the multi-magnification information of histopathological images. Furthermore, these visualization results illustrate the

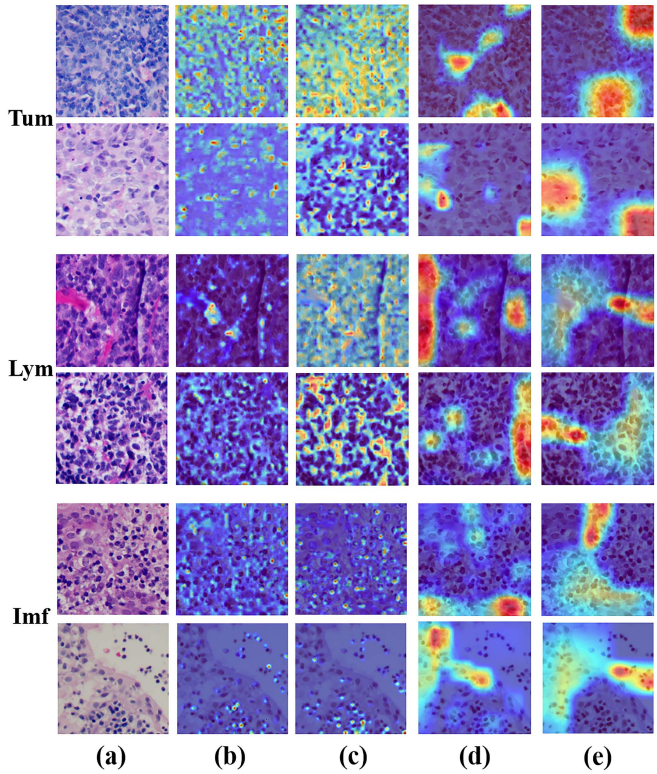

Fig. 5. Illustration of multi-magnification information exploited by the proposed DMSL method in NPC2020 dataset. (a) original histopathological image patches. (b), (c), (d), and (e) respectively give heat maps of (a) generated from the 5x of low dimension, 20x of low dimension, 5x of high dimension, and 20x of the high dimension of DMSL by utilizing a visualization technique called Grad-CAM. For the convenience of the display, the images were scaled to the same size.

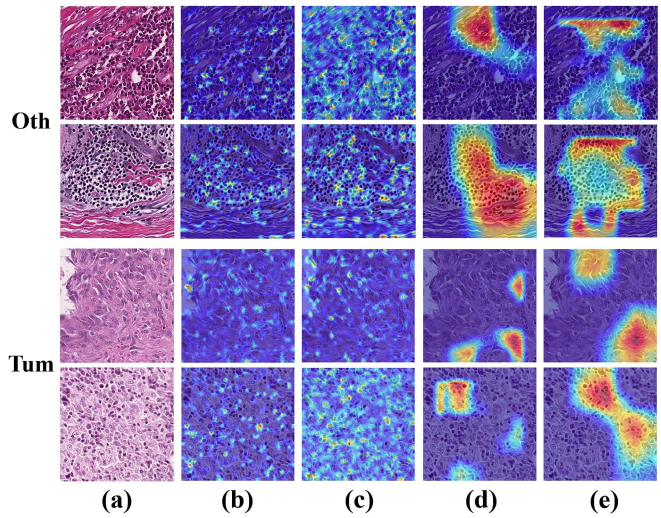

Fig. 6. Illustration of multi-magnification information exploited by the proposed DMSL method in BCSS2021 dataset. (a) original histopathological image patches. (b), (c), (d), and (e) respectively give heat maps of (a) generated from the 5x of low dimension, 20x of low dimension, 5x of high dimension, and 20x of the high dimension of DMSL by utilizing a visualization technique called Grad-CAM. For the convenience of the display, the images were scaled to the same size.

focus points of the model in different stages. The low-dimensional stage focuses more on cells. And in the high-dimensional stage, it focuses more on areas composed of multi-cellular and their microenvironment.

Specific to the performance of NPC2020 in Fig. 5, in the low dimensional stage of non-tumor regions, the feature regions of low magnification image learning were relatively concentrated. In contrast, the regions concerned by high magnification images were more scattered. Meanwhile, in the low dimensional stage of tumor regions, the feature of low magnification image learning was almost the same as that of high magnification image learning. DMSL focuses on larger areas and gives higher weights, no matter which category is in the high-dimensional stage. As for the BCSS2021 in Fig. 6, in the low dimensional stage of non-tumor, it was interesting that the features learned are very near to each other, whether at low or high magnification images, while the area learned at high magnification is still slightly larger. The distribution trend of the high-dimensional stage and two stages of the tumor was almost the same as that of the first dataset. Hence, it can be concluded that the model mainly learns the similarity information between the two images when learning the multi-magnification image, and the network can learn more abundant information in high magnification images than in low magnification images.

V. Conclusion

In summary, we propose a novel deep multi-magnification similarity learning approach that can be useful for the interpretation of multi-magnification learning framework and easy to visualize feature representation from low-dimension (e.g., cell-level) to high-dimension (e.g., tissue-level), which has overcome the difficulty of understanding cross-magnification information propagation. Moreover, we explore the reasons why multi-magnification image fusion is effective. This study showed that combining multi-magnification image information into the new artificial intelligence framework for the diagnosis of histopathological images is a more efficient and interpretable method. In this framework, the two images of different magnifications based on similarity are used to determine the classification results. The similarity is used to evaluate the synergy of the two images with different magnifications. We used two independent histopathological datasets to verify the proposed method, which was processed using the same program. Experimental results show that this method is effective and superior to the earlier methods.

However, with the continuous increase in the number of digital slides, it is important to further reduce the difficulty and time for pathologists to label annotations. This study is based on the image patches cropped from WSIs. In a future histopathological image classification task, we plan to develop a simpler and more efficient multi-magnification image directly applied to the framework of WSIs for diagnosis, gradually reducing the dependence on the pathologists' diagnostic process under the microscope and improving the calculation efficiency considering the current insufficient computing capacity. Therefore, in future work, we plan to: 1) achieve higher performance of the algorithm in histopathological diagnosis; 2) Combine algorithms with clinical devices such as microscopes to achieve application value. It will allow to promote the development of artificial intelligence in pathological diagnosis.

References

[1] F. Aeffner et al., "The gold standard paradox in digital image analysis: Manual versus automated scoring as ground truth," *Arch. Pathol. Lab. Med.,* vol. 141, no. 9, pp. 1267–1275, 2017.

[2] G. Yu et al., "Accurate recognition of colorectal cancer with semi-supervised deep learning on pathological images," *Nature Commun.,* vol. 12, no. 1, pp. 1–13, 2021.

[3] C.-L. Chen et al., "An annotation-free whole-slide training approach to pathological classification of lung cancer types using deep learning," *Nature Commun.,* vol. 12, no. 1, pp. 1–13, 2021.

[4] D. Li et al., "A deep learning diagnostic platform for diffuse large B-cell lymphoma with high accuracy across multiple hospitals," *Nature Commun.,* vol. 11, no. 1, pp. 1–9, 2020.

[5] N. Tokuyama et al., "Prediction of non-muscle invasive bladder cancer recurrence using machine learning of quantitative nuclear features," *Modern Pathol.*, vol. 35, no. 4, pp. 533–538, 2022.

[6] L. Wang et al., "Automated identification of malignancy in whole-slide pathological images: Identification of eyelid malignant melanoma in gigapixel pathological slides using deep learning," *Brit. J. Ophthalmol.,* vol. 104, no. 3, pp. 318–323, 2020.

[7] Z. Li, P. Zhang, N. Xie, G. Zhang, and C.-F. Wen, "A novel three-way decision method in a hybrid information system with images and its application in medical diagnosis," *Eng. Appl. Artif. Intell.,* vol. 92, 2020, Art. no. 103651.

[8] A. Belsare and M. Mushrif, "Histopathological image analysis using image processing techniques: An overview," *Signal Image Process.,* vol. 3, no. 4, pp. 23–36, 2012.

[9] O. Regnier-Coudert, J. McCall, R. Lothian, T. Lam, S. McClinton, and J. N'Dow, "Machine learning for improved pathological staging of prostate cancer: A performance comparison on a range of classifiers," *Artif. Intell. Med.*, vol. 55, no. 1, pp. 25–35, 2012.

[10] C. Atupelage et al., "Computational hepatocellular carcinoma tumor grading based on cell nuclei classification," *J. Med. Imag.*, vol. 1, no. 3, 2014, Art. no. 034501.

[11] V. Cheplygina, M. de Bruijne, and J. P. Pluim, "Not-so-supervised: A survey of semi-supervised, multi-instance, and transfer learning in medical image analysis," *Med. Image Anal.,* vol. 54, pp. 280–296, 2019.

[12] O. Jimenez-del-Toro et al., "Analysis of histopathology images: From traditional machine learning to deep learning," in *Biomedical Texture Analysis*. Amsterdam, The Netherlands: Elsevier, 2017, pp. 281–314.

[13] Z. Song et al., "Clinically applicable histopathological diagnosis system for gastric cancer detection using deep learning," *Nature Commun.,* vol. 11, no. 1, pp. 1–9, 2020.

[14] N. Coudray et al., "Classification and mutation prediction from non–small cell lung cancer histopathology images using deep learning," *Nature Med.,* vol. 24, no. 10, pp. 1559–1567, 2018.

[15] M. Gehrung, M. Crispin-Ortuzar, A. G. Berman, M. O'Donovan, R. C. Fitzgerald, and F. Markowetz, "Triage-driven diagnosis of Barrett's esophagus for early detection of esophageal adenocarcinoma using deep learning," *Nature Med.,* vol. 27, no. 5, pp. 833–841, 2021.

[16] C. Syrykh et al., "Accurate diagnosis of lymphoma on whole-slide histopathology images using deep learning," *NPJ Digit. Med.,* vol. 3, no. 1, pp. 1–8, 2020.

[17] M. Zaveri et al., "Recognizing magnification levels in microscopic snapshots," in *Proc. IEEE 42nd Annu. Int. Conf. Eng. Med. Biol. Soc.*, 2020, pp. 1416–1419.

[18] R. Schmitz et al., "Multi-scale fully convolutional neural networks for histopathology image segmentation: From nuclear aberrations to the global tissue architecture," *Med. Image Anal.,* vol. 70, 2021, Art. no. 101996.

[19] S. Kalra et al., "Pan-cancer diagnostic consensus through searching archival histopathology images using artificial intelligence," *NPJ Digit. Med.*, vol. 3, no. 1, pp. 1–15, 2020.

[20] F. Zhao et al., "Diagnosis of endometrium hyperplasia and screening of endometrial intraepithelial neoplasia in histopathological images using a global-to-local multi-scale convolutional neural network," *Comput. Methods Programs Biomed.*, vol. 221, 2022, Art. no. 106906.

[21] S. I. Khan, A. Shahrior, R. Karim, M. Hasan, and A. Rahman, "MultiNet: A deep neural network approach for detecting breast cancer through multi-scale feature fusion," *J. King Saud Univ.-Comput. Inf. Sci.,* vol. 34, no. 8, pp. 6217–6228, 2022.

[22] M. Nishio, M. Nishio, N. Jimbo, and K. Nakane, "Homology-based image processing for automatic classification of histopathological images of lung tissue," *Cancers,* vol. 13, no. 6, 2021, Art. no. 1192.

[23] C.-F. R. Chen, Q. Fan, and R. Panda, "Crossvit: Cross-attention multi-scale vision transformer for image classification," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2021, pp. 357–366.

[24] L. Tong, Y. Sha, and M. D. Wang, "Improving classification of breast cancer by utilizing the image pyramids of whole-slide imaging and multi-scale convolutional neural networks," in *Proc. IEEE 43rd Annu. Comput. Softw. Appl. Conf.*, vol. 1, 2019, pp. 696–703.

[25] K. Uehara, M. Murakawa, H. Nosato, and H. Sakanashi, "Multi-scale explainable feature learning for pathological image analysis using convolutional neural networks," in *Proc. IEEE Int. Conf. Image Process.*, 2020, pp. 1931–1935, doi: 10.1109/ICIP40778.2020.9190693.

[26] N. Marini et al., "Multi-scale task multiple instance learning for the classification of digital pathology images with global annotations," in *Proc. MICCAI Workshop Comput. Pathol.*, 2021, pp. 170–181.

[27] Y. Sun, X. Huang, Y. Wang, H. Zhou, and Q. Zhang, "Magnification-independent histopathological image classification with similarity-based multi-scale embeddings," 2021, *arXiv:2107.01063*.

[28] S. H. Gao, M.-M. Cheng, K. Zhao, X.-Y. Zhang, M.-H. Yang, and P. Torr, "Res2Net: A new multi-scale backbone architecture," *IEEE Trans. Pattern Anal. Mach. Intell.,* vol. 43, no. 2, pp. 652–662, Feb. 2021.

[29] Z. Zhou, X. Fan, P. Shi, and Y. Xin, "R-MSFM: Recurrent multi-scale feature modulation for monocular depth estimating," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2021, pp. 12777–12786.

[30] S. Doyle, M. Feldman, J. Tomaszewski, and A. Madabhushi, "A boosted Bayesian multiresolution classifier for prostate cancer detection from digitized needle biopsies," *IEEE Trans. Biomed. Eng.,* vol. 59, no. 5, pp. 1205–1218, May 2012.

[31] L.-C. Chen, Y. Yang, J. Wang, W. Xu, and A. L. Yuille, "Attention to scale: Scale-aware semantic image segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 3640–3649.

[32] K. Das, S. P. K. Karri, A. G. Roy, J. Chatterjee, and D. Sheet, "Classifying histopathology whole-slides using fusion of decisions from deep convolutional network on a collection of random multi-views at multi-magnification," in *Proc. IEEE 14th Int. Symp. Biomed. Imag.*, 2017, pp. 1024–1027.

[33] D. Gu et al., "Multi-scale patches convolutional neural network predicting the histological grade of hepatocellular carcinoma," in *Proc. IEEE 43rd Annu. Int. Conf. Eng. Med. Biol. Soc.*, 2021, pp. 2584–2587, doi: 10.1109/EMBC46164.2021.9630413.

[34] A. Tao, K. Sapra, and B. Catanzaro, "Hierarchical multi-scale attention for semantic segmentation," 2020, *arXiv:2005.10821*.

[35] T.-Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid networks for object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 2117–2125.

[36] S. Diao et al., "Weakly supervised framework for cancer region detection of hepatocellular carcinoma in whole-slide pathologic images based on multiscale attention convolutional neural network," *Amer. J. Pathol.,* vol. 192, no. 3, pp. 553–563, 2022.

[37] P. Viola and W. M. Wells, "Alignment by maximization of mutual information," *Int. J. Comput. Vis.*, vol. 24, pp. 137–154, 1997.

[38] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.,* vol. 13, no. 4, pp. 600–612, Apr. 2004.

[39] J. P. W. Pluim, J. B. A. Maintz, and M. A. Viergever, "Mutual-information-based registration of medical images: A survey," *IEEE Trans. Med. Imag.*, vol. 22, no. 8, pp. 986–1004, Aug. 2003.

[40] Y. Sun, X. Huang, H. Zhou, and Q. Zhang, "SRPN: Similarity-based region proposal networks for nuclei and cells detection in histology images," *Med. Image Anal.,* vol. 72, 2021, Art. no. 102142, doi: 10.1016/j.media.2021.102142.

[41] H.-T. Cheng et al., "Self-similarity student for partial label histopathology image segmentation," in *Proc. Eur. Conf. Comput. Vis.*, 2020, pp. 117–132.

[42] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 2818–2826.

[43] N. Marini et al., "Multi_Scale_Tools: A Python library to exploit multi-scale whole slide images," *Front. Comput. Sci.*, vol. 3, 2021, Art. no. 684521, doi: 10.3389/fcomp.2021.684521.

[44] J. Lin et al., "PDBL: Improving histopathological tissue classification with plug-and-play pyramidal deep-broad learning," *IEEE Trans. Med. Imag.,* vol. 41, no. 9, pp. 2252–2262, Sep. 2022.

[45] Z. Chen, J. Zhang, S. Che, J. Huang, X. Han, and Y. Yuan, "Diagnose like a pathologist: Weakly-supervised pathologist-tree network for slide-level immunohistochemical scoring," in *Proc. AAAI Conf. Artif. Intell.*, vol. 35, no. 1, 2021, pp. 47–54.

[46] Z. Tao, C. XiaoYu, L. HuiLing, Y. XinYu, L. YunCan, and Z. XiaoMin, "Pooling operations in deep learning: From 'invariable' to 'variable'," *BioMed Res. Int.,* vol. 2022, 2022, Art. no. 4067581.

[47] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Trans. Knowl. Data Eng.*, vol. 22, no. 10, pp. 1345–1359, Oct. 2010.

[48] V. Iglovikov and A. Shvets, "TernausNet: U-net with VGG11 encoder pre-trained on imageNet for image segmentation," 2018, *arXiv:1801.05746*.

[49] L. Wang, Y. Jiao, Y. Qiao, N. Zeng, and R. Yu, "A novel approach combined transfer learning and deep learning to predict TMB from histology image," *Pattern Recognit. Lett.,* vol. 135, pp. 244–248, 2020.

[50] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-CAM: Visual explanations from deep networks via gradient-based localization," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2017, pp. 618–626.

[51] A. Glowacz, "Fault diagnosis of electric impact drills using thermal imaging," *Measurement*, vol. 171, 2021, Art. no. 108815.

[52] A. Glowacz, "Ventilation diagnosis of angle grinder using thermal imaging," *Sensors,* vol. 21, no. 8, 2021, Art. no. 2853.

[53] X. Glorot, A. Bordes, and Y. Bengio, "Deep sparse rectifier neural networks," in *Proc. 14th Int. Conf. Artif. Intell. Statist.*, Fort Lauderdale, FL, USA, Apr. 11–13, 2011, vol. 15, pp. 315–323.

[54] S. Diao et al., "Computer-aided pathologic diagnosis of nasopharyngeal carcinoma based on deep learning," *Amer. J. Pathol.,* vol. 190, no. 8, pp. 1691–1700, 2020.

[55] M. Amgad et al., "Structured crowdsourcing enables convolutional segmentation of histology images," *Bioinformatics*, vol. 35, no. 18, pp. 3461–3467, 2019, doi: 10.1093/bioinformatics/btz083.

[56] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*.

[57] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 770–778.

[58] N. Ma, X. Zhang, H.-T. Zheng, and J. Sun, "ShuffleNet V2: Practical guidelines for efficient CNN architecture design," in *Proc. Eur. Conf. Comput. Vis.*, 2018, pp. 116–131.

[59] M. Tan et al., "MnasNet: Platform-aware neural architecture search for mobile," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 2820–2828.

[60] I. Radosavovic, R. P. Kosaraju, R. Girshick, K. He, and P. Dollár, "Designing network design spaces," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2020, pp. 10428–10436.

[61] W.-C. Huang et al., "Automatic HCC detection using convolutional network with multi-magnification input images," in *Proc. IEEE Int. Conf. Artif. Intell. Circuits Syst.*, 2019, pp. 194–198.

[62] B. Han, M. Zhang, X. Gao, Z. Wang, F. You, and H. Li, "Automatic classification method of thyroid pathological images using multiple magnification factors," *Neurocomputing,* vol. 460, pp. 231–242, 2021.

[63] F. M. Dickey and L. A. Romero, "Normalized correlation for pattern recognition," *Opt. Lett.*, vol. 16, no. 15, pp. 1186–1188, 1991.