UDC 004.75

C. Zhao

V. M. Dubovoy

Y. A. Horchuk

CLASSIFICATION METHOD FOR SKIN PATHOLOGICAL IMAGES

Vinnytsia National Technical University

Abstract

Skin pathological images contain essential diagnostic information across various scales. To effectively utilize multi-scale features, this study proposes a classification method based on multi-scale neural networks. The method involves a variable multi-scale neural network structure with a backbone network and multiple scale input branches inserted at different layers, facilitating feature extraction and fusion. Two search algorithms—a minimum cost-based search algorithm and a hill-climbing search algorithm—are introduced to identify the optimal network structure. Experimental results demonstrate that the proposed multi-scale network outperforms original networks in skin pathological image classification and that both search algorithms efficiently find near-optimal structures with reduced computational costs.

Keywords: skin pathology, multi-scale neural network, AMSICNN, deep learning, melanoma classification, image fusion, ResNet50, EfficientNetB0, InceptionV4, multi-scale input, CNN optimization, medical image analysis, hill-climbing search, minimum-cost search, whole-slide images.

1.Intelligent Data Analysis

The pathological diagnosis of skin pathological images typically requires pathologists with certain qualifications and extensive experience. When reviewing skin pathological images, pathologists first observe the overall situation of the sections in the low magnification mode of the microscope, identify suspicious lesion areas, and then use the high magnification mode to examine the morphological characteristics of each cell in the suspected lesion area and specific cell densities, usually requiring multiple repetitions of the above process until a diagnosis is reached. Obviously, facing a large number of undiagnosed skin pathological images, this manual diagnostic method is time-consuming, labor-intensive, and research has shown that there is a 25% diagnostic inconsistency among histopathology experts in distinguishing benign nevi from malignant melanomas[1]. Using computer-aided means to assist doctors in examining and diagnosing skin images can save a lot of manpower, material resources, and financial resources, and improve diagnostic efficiency. In recent years, with the powerful capabilities of deep learning, especially convolutional neural networks, in the field of computer vision, deep learning-based medical image processing has become a hot topic. Many research works on intelligent diagnosis of skin pathological images have emerged[2]. In the task of pathological image prediction, deep learning methods have achieved performance far beyond traditional methods.

However, these deep learning methods often directly adopt CNN models for identification, but these algorithms do not consider the characteristics of skin pathological images. In skin pathological image diagnosis, skin pathological images have multi-semantics, and key diagnostic information such as tumor size and extent, histological type, depth of infiltration, mitotic activity, margin status, presence of microsatellite or satellite metastases differ significantly in different magnification scales. Among them, features such as tumor size and extent, histological type, depth of infiltration, etc., need to be analyzed based on the overall situation of pathological tissues, and therefore these features are significant in low magnification pathological images. Features such as cell mitotic activity, margin status, presence of microsatellite or satellite metastases, etc., require detailed observation of lesion cells and immune cells, and therefore these features are significant in high magnification pathological images.

Addressing the problem of how to effectively extract and fuse pathological features at different scales, a skin pathological image classification method based on multi-scale neural networks is proposed. This method designs a variable multi-scale neural network structure and corresponding two multi-scale network structure search algorithms. The variable multi-scale neural network consists of a backbone network and parallel inserted multi-scale image input branches. The parallel input of multiple scale images enables the network to extract and fuse multi-scale image features. Based on the minimum cost-based search algorithm, this algorithm tests the influence of inserting input branches at different positions on the current network performance by specifying the priority of inserting different branch positions, discards input branches and insertion positions that reduce network performance, and selects favorable input branches to insert into the network. The hill-climbing search algorithm constructs a search space with all possible insertion positions, selects the most favorable position to insert input branches from the current search space each time, discards insertion positions that reduce network performance, until adding input branches cannot improve network performance or the search space is empty. Experimental results with ResNet50[3], EfficientnetB0[4], and InceptionV4[5] as backbone networks show that the multi-scale neural network achieves 0.4%~2.7% higher accuracy than the original network. Compared with the exhaustive method, the two multi-scale neural network search algorithms can find multi-scale network models close to the optimal solution at lower computational costs. Finally, the results of the ablation experiments indicate that the performance gain of the variable multi-scale neural network comes from additional image inputs.

2. Solving the Task

2.1 Machine learning model

To enhance the network's ability to extract multi-scale features, a novel multi-scale input neural network structure, called Alterable Multi-Scale Input Convolutional Neural Network (AMSICNN), is proposed. In a multi-scale input network, multiple-scale images are used as inputs, and the results are fused in the convolutional layers of the model. By learning from images of different scales, the model gains the ability to extract deep features of different scales. The multi-scale feature input network uses single-scale images as input and extracts features from different abstraction levels of convolutional modules, merging them into other layers to achieve multi-scale feature fusion. By combining multi-scale input networks and multi-scale feature input networks, the proposed AMSICNN achieves multi-scale feature extraction and fusion by inputting additional image information into different layers of the network.

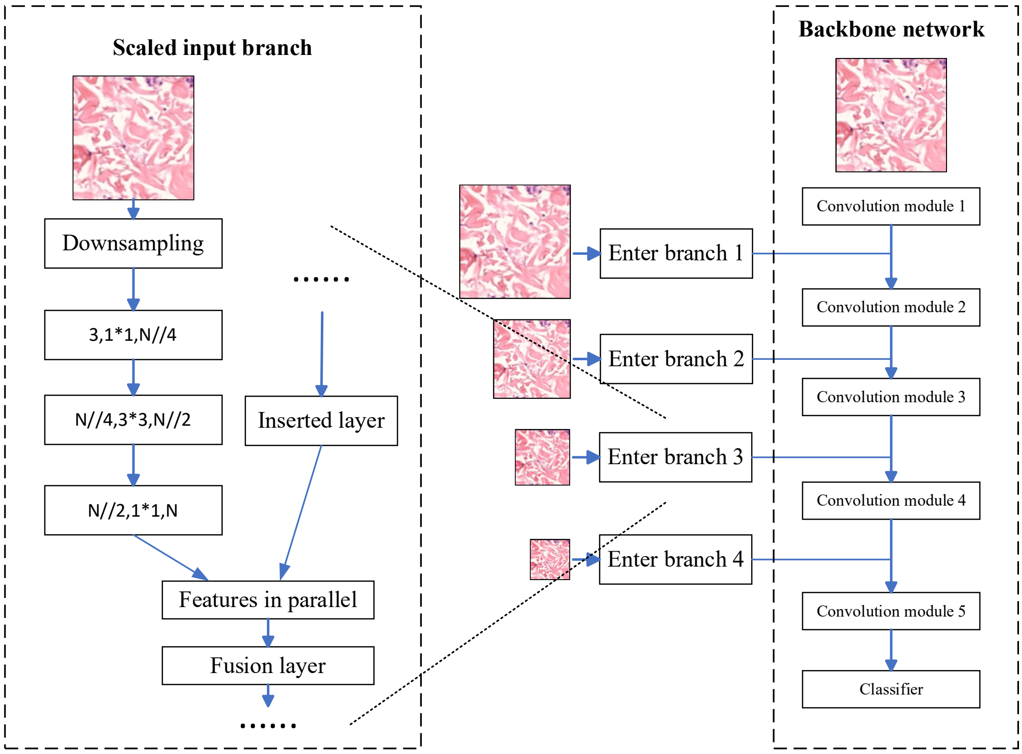

Fig. 1 Schematic diagram of variable multi-scale neural network

As shown in Figure 1, AMSICNN consists of a backbone network and multiple scale input branches. The multi-scale input branches are inserted into the backbone network in a parallel manner, and the output features of the branches are combined with the output features of the layer they are inserted into and then input into the next convolutional layer (fusion layer) of the backbone network. In the backbone network, input branches are only allowed to be inserted after downsampling layers, and the number of input branches inserted into AMSICNN can be changed. By adjusting the insertion positions and the number of input branches, AMSICNN can better adapt to different pathological image recognition tasks.

As illustrated in Figure 1, each multi-scale input branch consists of a downsampling layer and three convolutional layers, with convolutional kernel sizes of 1×1, 3×3, and 1×1, respectively. The downsampling layer in the input branch uses bilinear interpolation sampling, and its sampling rate is determined by the insertion position. The image is sampled to the same size as the output features of the layer it is inserted into. The channel numbers of the three convolutional layers in the input branch are determined by the number of channels of the layer they are inserted into, with output channel numbers of N/4, N/2, and N, respectively, where N is the number of channels of the layer they are inserted into. After processing by the downsampling layer and the three convolutional layers, the input image is expanded into a feature vector of the same size as the output features of the layer it is inserted into. The output features of the branch are concatenated with the output features of the layer they are inserted into and input into the fusion layer. The input channels of the fusion layer are doubled to accommodate the insertion of the branch.

As a new network structure composed of a backbone network and input branches, the number and insertion positions of the input branches determine the network's ability to extract and fuse multi-scale features. In AMSICNN, input branches can only be inserted into convolutional layers after downsampling layers in the backbone network. Assuming there are M insertion positions in the backbone network, theoretically, 2^M - 1 AMSICNN structures can be generated. AMSICNN has great flexibility in design, allowing for the selection of the optimal AMSICNN structure according to different tasks.

2.2 Network Search Algorithm

For a backbone network with M insertion positions, $2^m - 1$ AMSICNN structures can be generated. For AMSICNN, the impact of input branches on the network varies at different insertion positions, and when multiple input branches are inserted, the interaction between input branches can have additional effects on the network's performance.

The influence of multiple input branches on the network is not a simple linear superposition of the influence of a single input branch on the network. There is a possibility that inserting input branches may lead to a decrease in network performance. These characteristics of AMSICNN make it impossible to obtain the optimal AMSICNN structure through simple calculations. Instead, exhaustive methods are needed to train and test all AMSICNN networks, requiring a significant amount of computational resources. To address the problem of searching for the optimal AMSICNN with smaller computational costs, two heuristic search algorithms are proposed in this section:

(1) Minimum Cost-Based Multi-Scale Network Search Algorithm;

(2) Hill-Climbing Multi-Scale Network Search Algorithm, aiming to quickly find the optimal network structure.

2.2.1 Network Search Algorithm Based on Minimal Cost

To achieve an optimal multi-scale network structure with minimal computational cost, a multi-scale network search algorithm based on minimal cost is proposed. The algorithm begins with a backbone network and considers all potential insertion positions as the search space. The priority of each potential insertion position is determined by its depth in the network, with shallower positions having higher priority. The algorithm tests the potential insertion positions in order of priority. If inserting an input branch at a given position improves the current network performance, the input branch is permanently inserted into the network, and the combined network is used for the next step of the search. Otherwise, the insertion position is discarded. In this algorithm, each insertion position is tested only once, resulting in a computational complexity of $\mathrm{O(M \times T)}$, where M is the number of insertion positions and T is the computational cost of training and testing the model once. The detailed process of the algorithm is presented in Algorithm 1.

Given a selected backbone network $f_0(.)$ and dataset W ,all potential insertion positions q in$f_0(.)$constitute the search space $Q = \{q_1, q_2, \dots, q_m\}$, where m is the number of potential insertion positions in $f_0(.)$. The positions q are sorted by their depth in the network from shallow to deep. The specific steps of the algorithm are as follows:

1. The optimal network is initialized as the backbone network $f_{\text{best}}(.) = f_0(.)$ with the highest accuracy being the accuracy of the backbone network $\text{Acc}_{\text{best}} = f_0(W)$, i.e., the test result of the initial network $f_0(.)$after training on the dataset W .

2. Test the impact of an input branch at the i-th insertion position on model performance: Insert an input branch at position $q_i$in the optimal model $f_{\text{best}}(.)$to form the test model $f_{(i)}(.)$:

$$f_{(i)}(.) = f_{\text{best}}(.) + f_{\text{branch}}(q_i)(.) \tag{1}$$

Train and test the test model $f_{(i)}(.)$on the dataset W to obtain the test model accuracy:

$$\text{Acc}_i = f_{(i)}(W) \tag{2}$$

3. Compare the accuracy of the test model with that of the current optimal model. If the test model performs better than the current optimal model, update the optimal model and the highest accuracy; otherwise, the current optimal model remains unchanged.

4. Repeat steps (2) and (3) until the search space has been fully explored.

5. Output the optimal model $f_{\text{best}}(.)$.

This approach ensures that the optimal network structure is found with minimal computational expense by prioritizing and testing each potential insertion position only once.

2.2.2 Search Algorithm Based on Hill Climbing

Inspired by the hill climbing approach, a multi-scale network search algorithm based on hill climbing is proposed. Starting with a backbone network and constructing the search space from all possible insertion positions, the basic idea of the algorithm is to find the optimal insertion position q_k in the current search space Q , insert an input branch at that position, discard input branches that degrade network performance, and repeat the process until adding input branches no longer improves network performance or the search space is empty. The detailed process of the algorithm is presented in Algorithm 2.

The detailed steps of the algorithm are as follows:

1. Initialize the optimal network as the backbone network $f_{\text{best}}(.) = f_0(.)$ with the highest accuracy being the accuracy of the backbone network $\text{Acc}_{\text{best}} = f_0(W)$.

2. Test the impact of input branches at all insertion positions q in Q on the performance of the current optimal model: Insert an input branch at position q_i into the optimal model $f_{\text{best}}(.)$to form the test model $f_{(i)}(.)$:

$$f_{(i)}(.) = f_{\text{best}}(.) + f_{\text{branch}}(q_i)(.) \tag{3}$$

Train and test the test model $f_{(i)}(.)$on the dataset W to obtain the test model accuracy:

$$\text{Acc}_i = f_{(i)}(W) \tag{4}$$

3. Update the optimal model $f_{\text{best}}(.)$, highest accuracy, and search space based on the test model accuracies. Select the test model with the greatest performance improvement as the optimal model for the next search round, with its accuracy as the highest accuracy. If no test model shows performance improvement, the optimal model remains unchanged. The new search space consists of all insertion positions q_k that can improve performance, excluding the optimal insertion position q_j.

4. Repeat steps (2) and (3) until Q is empty.

5. Output the optimal model $f_{\text{best}}(.)$.

This method ensures that the optimal network structure is found efficiently by iteratively refining the search space and focusing on positions that enhance network performance.

3.Data in the dataset

The dataset used for validating the AMSICNN model and search algorithms is a multi-center melanoma pathology image dataset. It includes 1642 H&E-stained whole-slide images (WSIs) collected from three sources:

(1) Xiangya Hospital of Central South University (CSUXH)

Melanoma: 239 WSIs

Compound Nevus: 199 WSIs

Junctional Nevus: 169 WSIs

Intradermal Nevus: 188 WSIs

(2) The Cancer Genome Atlas (TCGA)

Melanoma: 22 WSIs

(3) Yale School of Medicine Tissue Microarray Center (YSM)

Melanoma: 825 WSIs

Total: 1642 WSIs

The original WSIs are extremely large (usually >100,000×100,000 pixels) and cannot be directly input into the CNN model for inference. The pathological images are preprocessed using a parallel method to handle WSIs and generate datasets. The WSIs, at 40× magnification, are processed into 512×512 pixel image patches using a sliding window method. All image patches are standardized, with blank background patches and non-lesion patches being discarded. Finally, the WSI images are divided into training, testing, and validation sets in a 7:1.5:1.5 ratio. The validation set is used for hyperparameter tuning, while the training and testing sets are used for neural network training and testing. Due to significant differences in the number of image patches among the four categories of melanoma, compound nevus, junctional nevus, and intradermal nevus, a certain number of image patches are randomly discarded or augmented from each WSI to achieve better CNN training results while ensuring the diversity of image patches. Notably, image patches at 20×, 10×, and 5× magnifications are downsampled from 40× magnification patches.

Table 1 Multi-Center Dataset

Data Source	Disease Type	Number of WSIs
CSUXH	Melanoma	239
	Compound Nevus	199
	Junctional Nevus	169
	Intradermal Nevus	188
TCGA	Melanoma	22
YSM	Melanoma	825
Total	1642	

5. Exploratory Data Analysis

The performance metrics used to evaluate the melanoma pathology image diagnosis model include accuracy (Acc), specificity, sensitivity, and F1 score.

Accuracy (Acc) is the most commonly used metric, representing the ratio of correctly classified image patches to all image patches. It indicates the model's ability to make correct diagnoses and can be used to evaluate the overall performance of the method. It can be expressed as:

$$Acc = \frac{N_{tp}+N_{tn}}{N_{tp}+N_{tn}+N_{fp}+N_{fn}} \tag{5}$$

Specificity and sensitivity are common features in medical diagnoses. Specificity refers to the probability of the diagnostic model not giving false positives, while sensitivity refers to the probability of not missing positive cases during diagnosis. The F1 score, which considers both precision and recall, is a commonly used evaluation metric for multi-class problems. It can be viewed as a harmonic mean of precision and recall.

To validate the classification ability of the AMSICNN model, ResNet50, VGG19, and EfficientNetB0 are selected as backbone networks. Tests are conducted on the original backbone network, AMSICNN (with all branches inserted), and AMSICNN (optimal structure). Table 3 shows the performance of all models in the four-class image patch classification task. AMSICNN (with all branches) refers to the AMSICNN model with all input branches inserted into the backbone network, while AMSICNN (optimal structure) refers to the optimal AMSICNN network model under the current backbone network.

As shown in Table 3, AMSICNN (optimal structure) achieves better performance than the original network across all three backbone networks. The improvement is most significant when using InceptionV4 as the backbone network, with a 2.6% increase in accuracy (Acc: 0.916 to 0.942). The improvement is less noticeable with EfficientNetB0 as the backbone network (Acc: 0.963 to 0.967). On the other hand, when ResNet50 is the backbone network, the F1 score of AMSICNN (all branches) is 0.951, and the F1 score of AMSICNN (optimal structure) is 0.953, showing close performance. However, when using EfficientNetB0 or InceptionV4 as the backbone network, the performance of AMSICNN (all branches) is significantly lower than that of AMSICNN (optimal structure), with a gap of up to 5.7% in accuracy for InceptionV4.

The experimental results indicate that the proposed AMSICNN model can classify melanoma and various nevi effectively. The additional scale information input of the AMSICNN network can enhance the diagnosis of melanoma. However, the performance improvement of the model with additional scale image input depends on the structure of the backbone network. The AMSICNN model's performance is 0.3% to 2.7% higher than that of the original network in all three backbone networks. Furthermore, the impact of input branch increase on model performance is non-linear, necessitating a multi-scale model search algorithm to identify the optimal network structure.

Table 3 Performance Comparison of AMSICNN and Original Models

Backbone Network	Model Structure	Accuracy	Sensitivity	Specificity	F1 Score
ResNet50	Original Network	0.956	0.929	0.972	**0.933**
	AMSICNN (All Branches)	0.967	0.951	0.979	**0.951**
	AMSICNN (Optimal Structure)|	0.969	0.956	0.978	**0.953**
EfficientNetB0	Original Network	0.963	0.942	0.977	**0.944**
	AMSICNN (All Branches)	0.96	0.932	0.977	**0.94**
	AMSICNN (Optimal Structure)|	0.967	0.949	0.948	**0.949**
InceptionV4	Original Network	0.916	0.832	0.977	**0.883**
	AMSICNN (All Branches)	0.885	0.832	0.953	**0.857**
	AMSICNN (Optimal Structure)	0.942	0.91	0.968	**0.918**

6. Analysis of the results

In this study, we addressed the challenge of multiscale feature extraction and fusion in pathological images by proposing a variable multiscale neural network architecture and two corresponding multiscale network structure search algorithms. The variable multiscale neural network consists of a backbone network and one or more parallel multiscale input branches. The parallel input of multiscale image information endows the network with the ability to extract and fuse multiscale features effectively. To identify the optimal multiscale network structure, we introduced two search algorithms: a minimum-cost search algorithm and a hill-climbing search algorithm. The minimum-cost search algorithm is designed to find the optimal structure with fixed and minimal computational cost, while the hill-climbing search algorithm seeks superior multiscale network structures at the expense of slightly higher computational costs. The experimental results demonstrated that the variable multiscale neural network outperformed the original network in diagnostic performance. Both search algorithms achieved near-optimal variable multiscale neural network structures with significantly lower computational costs compared to exhaustive search methods. Specifically, the minimum-cost search algorithm consistently required the least computational cost. In comparison, the hill-climbing search algorithm yielded more optimal multiscale network structures. An ablation study was conducted to further investigate the performance improvements of the variable multiscale network. The results indicated that the enhanced performance was primarily attributed to the additional multiscale image information input. This finding underscores the importance of incorporating multiscale inputs in neural networks for pathological image analysis.

In conclusion, the proposed variable multiscale neural network and the efficient search algorithms provide a promising approach for improving diagnostic performance in pathological image analysis while maintaining low computational costs.

References

1.Hekler A, Utikal J S, Enk A H, et al. Pathologist-level classification of histopathological melanoma images with deep neural networks[J]. Eur J Cancer, 2019,115: 79~83.

2.Wang L, Ding L, Liu Z, et al. Automated identification of malignancy in whole-slide pathological images: identification of eyelid malignant melanoma in gigapixel pathological slides using deep learning[J]. British Journal of Ophthalmology, 2020,104(3): 318~323.

3.He K, Zhang X, Ren S, et al. Deep residual learning for image recognition[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016:770~778

4.Tan M, Le Q V. EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks[J]., 2019.

5.Szegedy C, Ioffe S, Vanhoucke V, et al. Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning[J]., 2016.

Horchuk Yurii Anatoliyovych — postgraduate student of group 174-23a, faculty of computer systems and automation, Vinnytsia National Technical University, Vinnytsia, e-mail: yurii.horchuk@gmail.com

Caifeng Zhao — postgraduate student, faculty of computer systems and automation, Vinnytsia National Technical University, Vinnytsia, e-mail: zhaocaifeng0823@163.com

Volodymyr Dubovoy Mykhailovych — Dr. Sc. (Eng.), Professor, Department of Computer Control Systems, Vinnytsia National Technical University, Vinnytsia, e-mail: v.m.dubovoy@vntu.edu.ua

Ю. А. Горчук

В. М. Дубовой

Ц. Чжао

МЕТОД КЛАСИФІКАЦІЇ ПАТОЛОГІЧНИХ ЗОБРАЖЕНЬ ШКІРИ

Вінницький Національний Технічний Університет

Анотація

Патологічні зображення шкіри містять важливу діагностичну інформацію за різними шкалами. Для ефективного використання різномасштабних ознак у цьому дослідженні запропоновано метод класифікації на основі різномасштабних нейронних мереж. Метод включає змінну структуру багатомасштабної нейронної мережі з магістральною мережею та багатомасштабними вхідними гілками, вставленими на різних рівнях, що полегшує вилучення та злиття ознак. Два алгоритми пошуку - алгоритм пошуку на основі мінімальної вартості та алгоритм пошуку на основі сходження на гору - використовуються для визначення оптимальної структури мережі. Експериментальні результати показують, що запропонована багатомасштабна мережа перевершує оригінальні мережі в класифікації патологічних зображень шкіри і що обидва алгоритми пошуку ефективно знаходять близькі до оптимальних структури зі зменшеними обчислювальними витратами.

Ключові слова: патологія шкіри, багатомасштабна нейронна мережа, AMSICNN, глибоке навчання, класифікація меланоми, злиття зображень, ResNet50, EfficientNetB0, InceptionV4, багатомасштабний вхід, CNN-оптимізація, аналіз медичних зображень, пошук по висоті, пошук за мінімальною вартістю, зображення на весь слайд.

Горчук Юрій Анатолійович — аспірант групи 174-23а, факультет комп’ютерних систем і автоматики, Вінницький Національний Технічний Університет, Вінниця, e-mail:yurii.horchuk@gmail.com

Цайфенг Чжао — аспірантка, факультет комп’ютерних систем і автоматики, Вінницький Національний Технічний Університет, Вінниця, e-mail: zhaocaifeng0823@163.com

Дубовой Володимир Михайлович — д-р техн. наук, професор, кафедри комп'ютерних систем управління, Вінницький Національний Технічний Університет, Вінниця, e-mail: v.m.dubovoy@vntu.edu.ua