2024 IEEE International Symposium on Biomedical Imaging (ISBI) | 979-8-3503-1333-8/24/$31.00 ©2024 IEEE | DOI: 10.1109/ISBI56570.2024.10635572

CLASSIFYING MELANOCYTIC NEVUS BY USING EXTRACTED DEPTH-ENCODED CHANNEL INFORMATION

Yaoqi Tang[1], Qilin Sun[2], Siqi Wang[1], Bo Jiang[3], Yuye Ling[,1]*

[1]Department of Electronic Engineering, Shanghai Jiao Tong University, Shanghai 200240, China
[2]Department of Dermatology, Ninth People's Hospital
Affiliated to Shanghai Jiao Tong University School of Medicine, Shanghai 200011, China
[3]John Hopcroft Center for Computer Science, Shanghai Jiao Tong University, Shanghai 200240, China

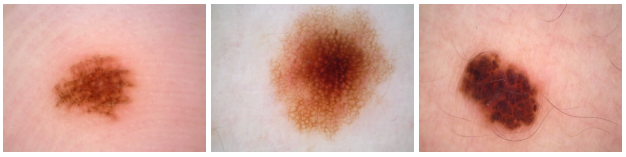
ABSTRACT

Melanocytic nevi are common benign skin lesions. However, some types of nevi have a higher risk of malignant transformation into melanoma, which is the most aggressive form of skin cancer. To avoid unnecessary biopsies and help reduce medical costs, accurate fine-grained nevi classification methods are important. Different from skin lesion classification, melanocytic nevi classification is more challenging due to the inter-class similarity. Inspired by the physical principles underlying dermoscopy imaging and pathological basis of nevi, we propose a novel classification network to extract depth-relevant channel-specific information and channel-shared information for dermoscopy images. A reconstruction module with a differential loss is used to exploit and enlarge the representations of different channels. Our algorithm is evaluated on a three-class melanocytic nevi dataset with ground truth verified by histological examination. The accuracy of the proposed method is 3.69% higher than state-of-the-art skin lesion classification technique.

Index Terms— Melanocytic Nevus, Image Classification, Color Channel, Depth

1. INTRODUCTION

Skin lesion refers to the area that has different characteristics from the surrounding skin. The majority of them are non-cancerous and harmless, such as melanocytic nevi, which are benign proliferations of melanocytes residing in the epidermis and/or dermis. However, despite their benignity, certain histological subtypes have a potential risk of malignant transformation and harbor oncogenic mutations associated with melanoma [1], which is the most dangerous type of skin cancer and causes about 55,500 cancer deaths worldwide annually [2]. A histological examination revealed that approximately 33% of melanomas are derived directly from benign melanocytic nevi [3]. Clinical features of these nevi often overlap, leading to reliance on histopathological classification. Unfortunately, blind biopsies impose a dual burden on patients and healthcare resources. Despite that a single overall mean visit cost for a skin lesion screening was $150, there are many unnecessary costs for screening and biopsy of benign types since the number needed to screen to detect one skin cancer was 12.2 and melanoma was up to 215 [4]. Thus, an effective early fine-grained detection of melanocytic nevi is crucial for preventive diagnosis.



(a) Junctional nevus. (b) Compound nevus. (c) Intradermal nevus.

Fig. 1: Example dermoscopy images of melanocytic nevi.

Conventionally, dermatologists analyze the appearance of suspicious rashes and determine whether follow-up or surgical excision is necessary based on dermoscopy images, which can provide deeper and clearer details visualization of the skin structure by utilizing polarized light [5]. Exemplary dermoscopy images of melanocytic nevi are shown in Fig 1. However, accurate clinical dermoscopic diagnosis is highly dependent upon the dermatologist's expertise [6]. Hence, it is necessary to establish an automated computerized classification system for melanocytic nevus to promote dermoscopy technology and to assist dermatologists in decision-making.

For automated classification, researchers in previous work have proposed various algorithms for benign and malignant skin lesion classification, which focus more on the differences in image texture features. In general, traditional methods first extract handcrafted features such as texture and shape on the dermoscopy image, followed by a classification scheme using several standard machine-learning approaches such as SVM and decision tree [7]. However, these approaches have demonstrated limited generalizability. Recently, various deep learning based techniques have been proposed. Zhang *et al.* [8] proposed an attention residual learning convolutional neu-

*Corresponding author: Yuye Ling (yuye.ling@sjtu.edu.cn)

ral network to improve discriminative representation. Yao *et al.* [9] proposed a single-model based cumulative learning strategy with a multi-weighted new loss for classification on small and imbalanced skin lesion datasets. Gessert *et al.* [10] proposed a patch-based attention model with diagnosis-guided loss weighting to solve the class imbalance problem and utilize the images more effectively. However, these methods typically focus on modifying modules, adopting other training strategies, or adjusting weighted loss to extract image information for benign/malign distinction, but do not explicitly consider the pathological differences across these skin lesions, especially for preventive diagnosis.

To address above issues, we make a connection between the physical principles of dermoscopy imaging and the pathological classification basis of melanocytic nevi. Inspired by the penetration depth of polarized light through skin, we consider that different color channels in polarized dermoscopy images contain special information within different depths of the skin, which could assist melanocytic nevi classification. Based on this assumption, we propose a classification model consisting of a channel-specific feature extraction branch and a channel-shared information fusion branch. The main contribution lies in the first branch, which includes new encoder modules and a differential loss to better separate and widen feature expressions between different channels. Moreover, a reconstruction module is implemented to constrain and regularize the channel-specific features. By fusing features from both branches, the performance is lifted compared to baseline and state-of-the-art skin lesion classification methods.

2. METHOD

Our proposed method consists of two branches: a channel-specific feature extraction branch and a channel-shared information fusion branch. The design of the first branch is motivated by the pathological basis of nevi and physical principles of dermoscopy imaging. The pathological classification basis of different nevi is the position of melanocyte nests. Based on the distribution of melanocytes observed by histopathological analysis, melanocytic nevi are divided into three kinds: junctional nevi(at the dermo-epidermal junction), intradermal nevi (within the dermis) and compound nevi (both) [11], as shown in Fig. 2. As a matter of fact, polarized light undergoes several scattering and absorption events that decrease the penetration depth in the skin tissue before being back-scattered to the detection device. In general, in the visible-light band, longer wavelengths of polarized light correspond to smaller scattering coefficients of the tissue, implying fewer scattering events [12]. This fact indicates that the red light may penetrate deeper within the skin compared to the green and blue light. Hence, it is reasonable to assume that some unique information of melanocytic nevi at varying depths is carried by different color channels. Inspired by this, we design a novel branch to extract unique information for different color channels. The overall framework of the proposed method is shown in Fig. 3.

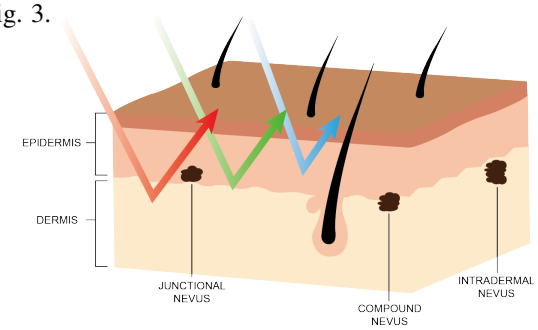


Fig. 2: The schematic penetrating and back-scattered light at different wavelengths with skin.

2.1. Channel-specific Feature Extraction Branch

Aiming at extracting specific channel information related to a different depth, we construct the channel-specific feature extraction branch based on four depthwise blocks, as shown in Fig. 3. Each depthwise block consists of a few stacked layers, including convolutional layers, batch normalization (BN) layers, and ReLU layers. In order to prevent information interference between different channels, we use separable convolutions to extract channel features. Specifically, the features are separated into three parts, which are then processed by two group convolutions to achieve feature extraction and dimensionality transformation. After stacking them back together, a simple channel attention mechanism [13] is used to assign better weights to more important channels. Additionally, skip connections are employed to maintain the input features.

After obtaining the independent channel-specific information from the encoder, we utilize a differential loss to control the training process. This loss helps to separate the learned representations of each channel, thus reducing their similarity and enabling each channel to have diverse feature representations. The differential loss is defined as follows:

$$L_{diff} = \sum_{i,j=1,i\neq j}^{3} \mathrm{sim}(\mathrm{z_i}, \mathrm{z_j}) \tag{1}$$

where z_i and z_j denote the final features extracted from the channel-specific feature extraction branch. And we use the cosine similarity as the sim function.

Besides, to better extract the input channel information and prevent the learned channel representations from going beyond a certain reasonable range, an additional reconstruction module is employed to constrain and regularize the learned features of the channel information. Since the information from three channels is derived from the same dermoscopy image, we used a unified upsampling architecture for reconstruction and calculated the mean squared error (MSE) loss with the original image. The reconstruction module contains three upsampling blocks and a final convolutional layer, each of which includes a set of convolution

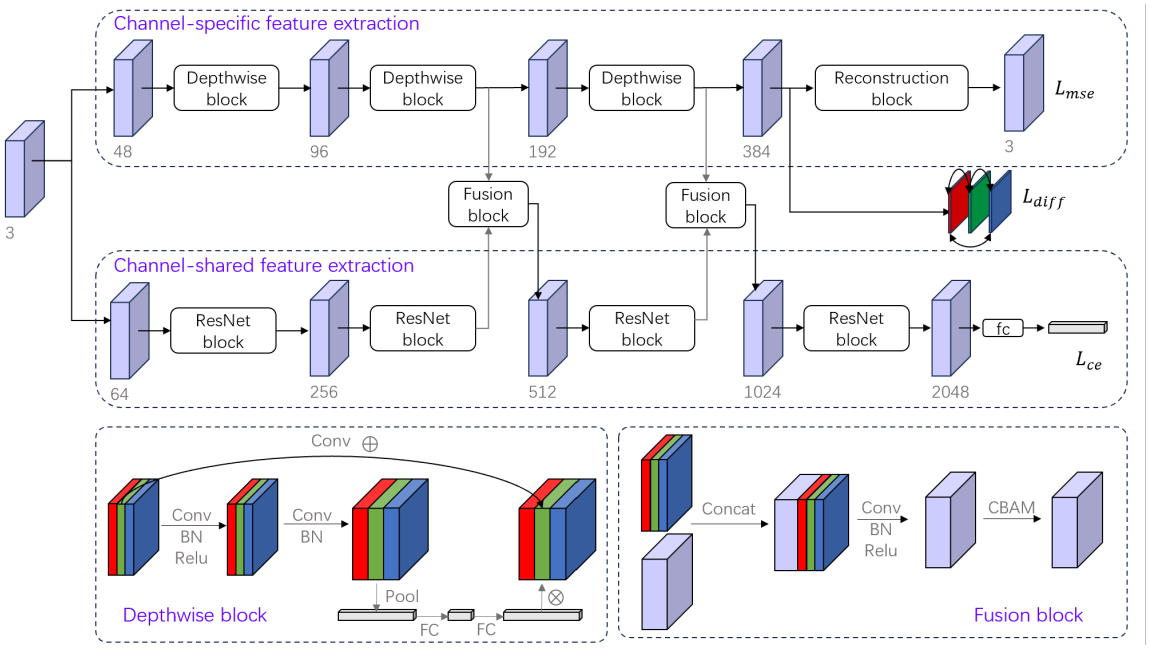


Fig. 3: The framework of the proposed method. It consists of two feature extraction branch fused through a fusion module.

layers and deconvolution layers. The loss function for the reconstruction module is defined as follows:

$$L_{mse} = \sum_{i=1}^{3} \|y_i^r - y_i\|_2 \tag{2}$$

where y_i and y_i^r denote the i_{th} channel for the input image and the reconstruction image.

2.2. Channel-shared Feature Extraction Branch

In this branch, the channel-shared information related to three classes from the channels is jointly extracted in a fusion way by a classic classification encoder. To make full use of the information from the upper branch, we propose a fusion block to fuse the specific information from the first branch to the shared information. As shown in Fig. 3, the fusion module concatenates features from two branches and then uses a convolutional layer to fuse the information. Since the features come from different branches and channels, we need to adjust the overall weights to normalize the features. Therefore, we adopt the CBAM [14] module to improve the weights of important channels and spatial features. Thus, by fusing features from both branches, the specific and shared information is extracted for the final classification. The output dimension of the fully connected (FC) layer is modified to the number of classes, in which each value presents a probability score to a certain type. Finally, we utilize cross-entropy loss L_{ce} to optimize the branch training and improve final classification accuracy. The overall loss for the end-to-end framework is defined as:

$$L = \alpha L_{diff} + \beta L_{mse} + L_{ce} \tag{3}$$

where α and β denotes the loss weights.

3. EXPERIMENTS

3.1. Dataset

We retrospectively analyzed the melanocytic nevi dataset containing 2,150 dermoscopic data from 1,634 subjects who underwent histopathological examination between January 2020 and December 2022 at the Ninth People's Hospital Affiliated to Shanghai Jiao Tong University School of Medicine. These dermoscopy images were taken by Taiwan CBS skin detector. Melanocytic nevi could be divided into three classes, with 316 images of junctional nevus, 612 images of compound nevus and 1,222 images of intradermal nevus. The labels of the dataset are the gold standard verified by histological examination. We randomly split the images of each type into a training subset and a test subset with a 9:1 ratio. A five-fold validation strategy is used for better evaluation.

3.2. Experimental Setting

We implement the proposed method using PyTorch and adopt ResNet50 [15] as the backbone of the channel-shared feature extraction branch with the initialized weights pretrained on the ImageNet dataset. All the training and testing processes were run on an NVIDIA 3090 GPU. The network is trained in an end-to-end manner using the adaptive moment estimation (Adam) optimizer with a mini-batch of 8 images for about 30 epochs. The learning rate is initialized to 0.0001 and trained from scratch with a 3 epoch warm-up.

For convergence, training images will be resized to 224×224. During the training phase, we conduct data augmentations, including random crop with a portion of 0.5 to 1, ran-

Table 1: The detailed classification results (%) of different methods. ACC, AUC are short for accuracy, and area under curve, respectively. The best performance is marked in red.

Method	Junctional nevus		Compound nevus		Intradermal nevus		ACC(avg)	AUC(avg)
	Precision	Recall	Precision	Recall	Precision	Recall		
ResNet50 [15]	49.67 ± 5.24	46.67 ± 6.60	38.33 ± 3.70	34.33 ± 2.94	76.87 ± 0.17	81.00 ± 1.41	64.96 ± 1.33	72.22 ± 0.61
Efficient-b4 [16]	53.96 ± 2.13	49.93 ± 2.36	37.91 ± 5.67	35.61 ± 5.24	76.51 ± 2.69	80.70 ± 2.82	65.43 ± 1.22	72.78 ± 1.14
Mbit [9]	57.78 ± 2.22	49.39 ± 2.61	40.06 ± 4.06	34.38 ± 3.13	77.08 ± 0.78	80.92 ± 2.29	65.82 ± 1.17	72.31 ± 0.45
TABE [17]	56.54 ± 3.03	48.15 ± 3.80	41.45 ± 2.32	40.83 ± 6.69	77.68 ± 1.99	81.02 ± 1.88	66.54 ± 1.08	73.30 ± 0.30
Proposed	59.75 ± 4.99	52.03 ± 3.69	49.93 ± 3.76	36.90 ± 2.40	77.70 ± 0.56	87.34 ± 1.33	70.23 ± 1.77	75.13 ± 0.73

dom rotation, flipping, color jittering. In particular, the rotation is performed from -10 to 10 degrees. The weights α and β for training are set to be $\frac{1}{3}$ and 1. The metrics chosen for evaluating the performance include accuracy (ACC), area under the ROC curve (AUC), precision and recall.

3.3. Benchmark Study

In order to verify the proposed framework, we compare the performance of our method with baseline and current state-of-the-art (SOTA) methods, using the code released by the authors. In detail, two baseline methods include ResNet50 [15] and EfficientNet [16], and SOTA methods include Mbit [9] and TABE [17] in skin lesion classification tasks are chosen.

The qualitative results of different methods are reported in TABLE 1. The experiments are repeated three times with different random seeds. It can be observed that our model achieves the highest precision, accuracy and AUC among all other methods. Compared to baseline methods such as ResNet50, our model achieves an improvement of 10.08% in precision of junctional nevus and 5.27% in average accuracy. Moreover, the accuracy is improved by 3.69% compared to Mbit, although the proposed method does not yield much improvement in recall for compound nevus classification. We also note that the performance of compound nevus is lower than that of junctional nevus, even though the number of junctional nevus is smaller. We suspect that this is because the compound nevus is not an independent class, which has information within the other two types. The results in TABLE 1 have shown that the change in model architecture or training strategy is not suitable for classifying such a nevus classification task with low inter-class variance.

3.4. Ablation Study

To evaluate the effectiveness of each module of the network, the ablation study consists of three main aspects: the impact of channel-specific feature extraction branch, the effect of reconstruction loss, and the influence of differential loss. The quantitative results of the ablation study on our dataset are listed in TABLE 2.

We first investigate the impact of the channel-specific branch for extracting depth-related channel-specific information. The baseline is the one-branch (OB) pretrained ResNet50 network with a modified fully-connected layer used in our channel-shared feature extraction branch. Two-branch (TB) baseline is the proposed method without the reconstruction module (RM) and the differential loss (DL). From TABLE 2, we could find that the accuracy of two-branch network is improved by 1.44% compared with that of OB baseline. The depth-related information of the adding branch is important in this task, since the three classes of nevus differ less in feature space. We then investigate the impact of the reconstruction module. The regularization of the reconstruction part makes the learned information more robust and reliable. From TABLE 2, after introducing the reconstruction module into the network, the AUC and accuracy are both lifted, achieving an improvement of 2.00% compared with that of the TB baseline. Moreover, the differential loss is important for channel-specific information extraction, since it prevents the training of the upper branch from becoming similar to the bottom branch.

Table 2: Comparison of different ablation study models in terms of accuracy (%) and AUC (macro).

Method	OB	TB	DL	RM	ACC	AUC
One-branch baseline	✓				65.77	81.95
Two-branch baseline		✓			67.21	82.68
Proposed w/o RM		✓	✓		68.56	83.19
Proposed w/o DL		✓		✓	69.21	83.24
Proposed		✓	✓	✓	70.23	83.57

4. CONCLUSION

In this paper, we propose a two-branch melanocytic nevi classification framework to extract depth-related information for dermoscopy images. Representations of different channels are extracted and constrained by a differential loss and a reconstruction module. Moreover, we utilize a fusion scheme to perform information fusion, which then undergoes a classification layer to obtain the final diagnosis. Our method is evaluated on a three-class melanocytic nevus dataset with ground truth verified by histological examination. Experiments and ablation studies have shown that our method achieves a higher accuracy compared to other methods.

5. ACKNOWLEDGMENT

The authors would like to thank Jianing Mao for his advice on skin tissue optics to this paper. This work is supported by

"the Fundamental Research Funds for the Central Universities (project number YG2022QN058)".

6. REFERENCES

[1] Mi Ryung Roh, Philip Eliades, Sameer Gupta, and Hensin Tsao, "Genetics of melanocytic nevi," *Pigment Cell & Melanoma Research*, vol. 28, no. 6, pp. 661–672, 2015.

[2] Dirk Schadendorf, Alexander CJ van Akkooi, Carola Berking, Klaus G Griewank, Ralf Gutzmer, Axel Hauschild, Andreas Stang, Alexander Roesch, and Selma Ugurel, "Melanoma," *The Lancet*, vol. 392, no. 10151, pp. 971–984, 2018.

[3] WE Damsky and Marcus Bosenberg, "Melanocytic nevi and melanoma: unraveling a complex relationship," *Oncogene*, vol. 36, no. 42, pp. 5771–5792, 2017.

[4] Martha Matsumoto, Aaron Secrest, Alyce Anderson, Melissa I. Saul, Jonhan Ho, John M. Kirkwood, and Laura K. Ferris, "Estimating the cost of skin cancer detection by dermatology providers in a large health care system," *Journal of the American Academy of Dermatology*, vol. 78, no. 4, pp. 701–709.e1, 2018.

[5] Michael Binder, Margot Schwarz, Alexander Winkler, Andreas Steiner, Alexandra Kaider, Klaus Wolff, and Hubert Pehamberger, "Epiluminescence Microscopy: A Useful Tool for the Diagnosis of Pigmented Skin Lesions for Formally Trained Dermatologists," *Archives of Dermatology*, vol. 131, no. 3, pp. 286–291, 03 1995.

[6] Mihaela Balu, Kristen M. Kelly, Christopher B. Zachary, Ronald M. Harris, Tatiana B. Krasieva, Karsten König, Anthony J. Durkin, and Bruce J. Tromberg, "Distinguishing between Benign and Malignant Melanocytic Nevi by In Vivo Multiphoton Microscopy," *Cancer Research*, vol. 74, no. 10, pp. 2688–2697, 05 2014.

[7] Md. Kamrul Hasan, Md. Asif Ahamad, Choon Hwai Yap, and Guang Yang, "A survey, review, and future trends of skin lesion segmentation and classification," *Computers in Biology and Medicine*, vol. 155, pp. 106624, 2023.

[8] Jianpeng Zhang, Yutong Xie, Yong Xia, and Chunhua Shen, "Attention residual learning for skin lesion classification," *IEEE transactions on medical imaging*, vol. 38, no. 9, pp. 2092–2103, 2019.

[9] Peng Yao, Shuwei Shen, Mengjuan Xu, Peng Liu, Fan Zhang, Jinyu Xing, Pengfei Shao, Benjamin Kaffenberger, and Ronald X Xu, "Single model deep learning on imbalanced small datasets for skin lesion classification," *IEEE transactions on medical imaging*, vol. 41, no. 5, pp. 1242–1254, 2021.

[10] Nils Gessert, Thilo Sentker, Frederic Madesta, Rüdiger Schmitz, Helge Kniep, Ivo Baltruschat, Rene Werner, and Alexander Schlaefer, "Skin lesion classification using cnns with patch-based attention and diagnosis-guided loss weighting," *IEEE Transactions on Biomedical Engineering*, vol. 67, no. 2, pp. 495–503, 2019.

[11] Giovanni Pellacani, Alon Scope, Barbara Ferrari, Gaia Pupelli, Sara Bassoli, Caterina Longo, Anna Maria Cesinaro, Giuseppe Argenziano, Rainer Hofmann-Wellenhof, Josep Malvehy, Ashfaq A. Marghoob, Susana Puig, Stefania Seidenari, H. Peter Soyer, and Iris Zalaudek, "New insights into nevogenesis: In vivo characterization and follow-up of melanocytic nevi by reflectance confocal microscopy," *Journal of the American Academy of Dermatology*, vol. 61, no. 6, pp. 1001–1013, 2009.

[12] Valery V Tuchin et al., "Tissue optics," Society of Photo-Optical Instrumentation Engineers (SPIE) Bellingham, WA, USA, 2015.

[13] Jie Hu, Li Shen, and Gang Sun, "Squeeze-and-excitation networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2018.

[14] Sanghyun Woo, Jongchan Park, Joon-Young Lee, and In So Kweon, "Cbam: Convolutional block attention module," in *Proceedings of the European Conference on Computer Vision (ECCV)*, September 2018.

[15] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2016.

[16] Mingxing Tan and Quoc Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," in *Proceedings of the 36th International Conference on Machine Learning*, Kamalika Chaudhuri and Ruslan Salakhutdinov, Eds. 09–15 Jun 2019, vol. 97 of *Proceedings of Machine Learning Research*, pp. 6105–6114, PMLR.

[17] Peter J. Bevan and Amir Atapour-Abarghouei, "Detecting melanoma fairly: Skin tone detection and debiasing for skin lesion classification," in *Domain Adaptation and Representation Transfer*, Konstantinos Kamnitsas, Lisa Koch, Mobarakol Islam, Ziyue Xu, Jorge Cardoso, Qi Dou, Nicola Rieke, and Sotirios Tsaftaris, Eds., Cham, 2022, pp. 1–11, Springer Nature Switzerland.