2021 43rd Annual International Conference of the
IEEE Engineering in Medicine & Biology Society (EMBC)
Oct 31 - Nov 4, 2021. Virtual Conference

Multi-level Attentive Skin Lesion Learning for Melanoma Classification

Xiaohong Wang[1], Weimin Huang[1], Zhongkang Lu[1], and Su Huang[1]

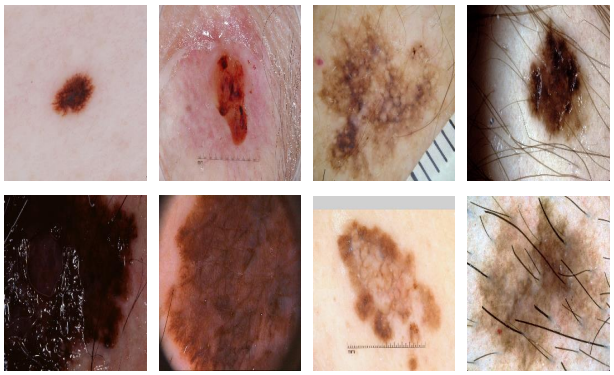
***Abstract*— Melanoma classification plays an important role in skin lesion diagnosis. Nevertheless, melanoma classification is a challenging task, due to the appearance variation of the skin lesions, and the interference of the noises from dermoscopic imaging. In this paper, we propose a multi-level attentive skin lesion learning (MASLL) network to enhance melanoma classification. Specifically, we design a local learning branch with a skin lesion localization (SLL) module to assist the network in learning the lesion features from the region of interest. In addition, we propose a weighted feature integration (WFI) module to fuse the lesion information from the global and local branches, which further enhances the feature discriminative capability of the skin lesions. Experimental results on ISIC 2017 dataset show the effectiveness of the proposed method on melanoma classification.**

***Index Terms*— melanoma classification, multi-level attentive skin lesion learning, skin lesion localization, weighted feature integration.**

I. INTRODUCTION

Melanoma is one of the most dangerous skin lesion diseases because of its uncontrolled growth and high potential to infect other parts of the body [1]–[4]. Early diagnosis of melanoma is crucial for the timely treatment of the skin lesions [5]–[8]. Recently, melanoma classification by deep learning has attracted a lot of research interests. For example, Yang *et al*. [9] used a convolution neural network based on region average pooling for skin lesion classification. Harangi [10] ensembled a set of deep learning networks to improve the performance of skin lesion classification. Song *et al*. [11] designed an end-to-end multi-task deep learning framework for skin lesion analysis. Sultana *et al*. [12] built a deep residual network with regularised fisher framework for melanoma detection. Al *et al*. [13] diagnosed multiple skin lesions via the integrated deep convolutional networks. Yu *et al*. [14] exploited the lesion localization information from the segmentation to improve the performance of melanoma classification. Zhang *et al*. [15] utilized an attention residual learning for skin lesion classification. Yu *et al*. [16] classified the demorscopic images represented by the local convolutional features extracted from a deep residual network. Gu *et al*. [17] used a two-step progressive transfer learning and adversarial domain adaption for skin lesion classification. Although those deep learning based methods have achieved promising performance for skin lesion classification, it is still a challenging task due to the appearance variation of skin lesion at different progression stages (Fig. 1 (a-b)), and the noises like hairs, air bubbles and ruler-like non-skin objects (Fig. 1 (c-d)).

[1]Xiaohong Wang, Weimin Huang, Zhongkang Lu, and Su Huang are with the Institute for Infocomm Research, A*STAR, Singapore.



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

Fig. 1. Examples of melanoma cases with different appearances.

In this paper, we propose a novel multi-level attentive skin lesion learning (MASLL) network for melanoma classification. To understand the lesion structures with different appearances, we propose to look into skin lesions from two branches, i.e., global branch and local branch. Specifically, we propose a local learning branch with a skin lesion localization (SLL) module to assist the network in learning lesion features from the region of interest. By embedding SLL into the deep feature learning process of skin lesions, we generate a specific lesion feature representation from the salient skin lesion regions, which further enhances the interpretability of the proposed network on skin lesion classification. In addition, to improve the feature parsing ability of the network on skin lesion classification, we propose a weighted feature integration module to fuse the lesion feature information from the global branch and the local branch. By exploiting the lesion feature information from two branches, we not only enrich the feature description of the skin lesion, but also reduce the effect of the noises on skin lesion classification. Our proposed network is an end-to-end deep architecture for melanoma classification. The main contributions of the proposed method are summarized as:

(a) We propose a local learning branch with a skin lesion localization module, which assists the network in concentrating on feature learning from the salient lesion regions.

(b) We propose a weighted feature integration module, which enriches the feature representation of skin lesion by fusing the feature information from global and local branches.

(c) Experimental analysis on publicly available dataset shows that the proposed network has achieved better melanoma classification performance than recent melanoma classification works.

978-1-7281-1178-0/21/$31.00 ©2021 IEEE

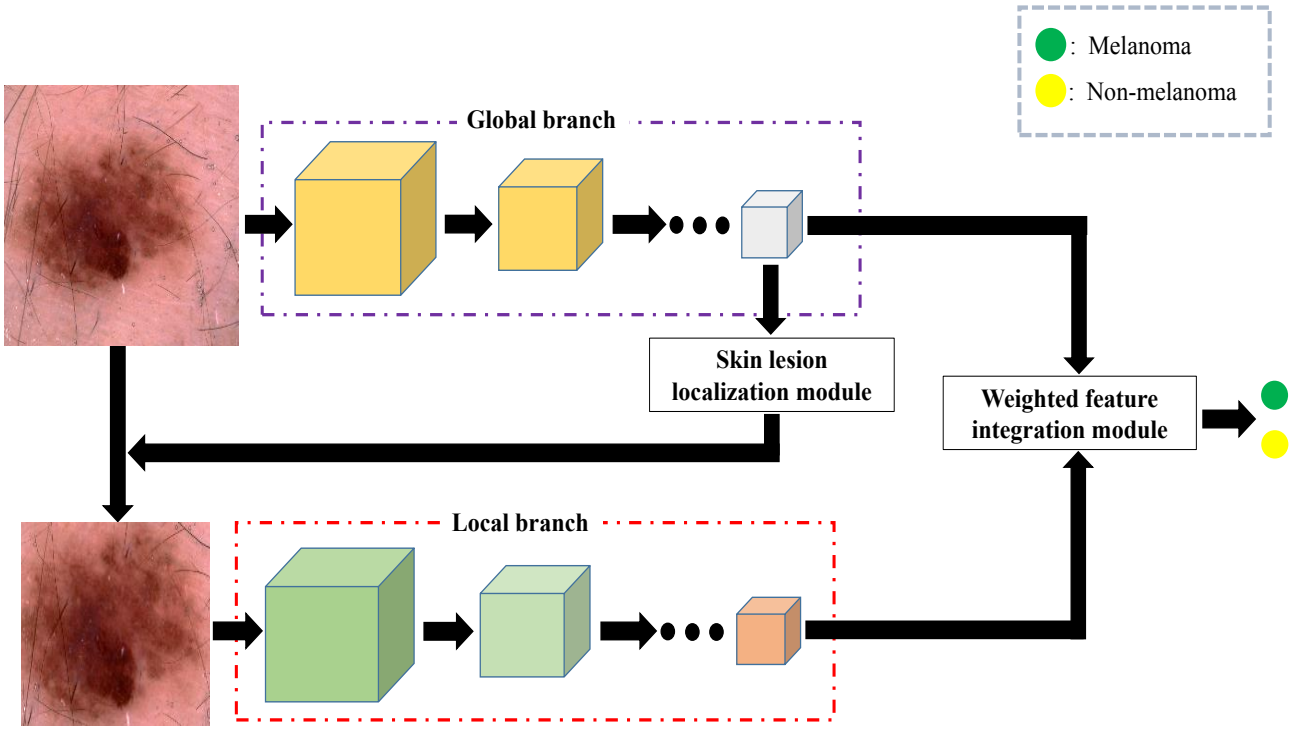


Fig. 2. Flowchart of the proposed multi-level attentive skin lesion learning network for melanoma classification.

II. Proposed Method

In this section, we describe the proposed skin lesion classification in detail. Fig. 2 shows the architecture of the proposed network. Our proposed network extracts the skin feature information from two branches, i.e., global branch and local branch. The global branch learns the lesion feature from the whole dermoscopic images, while the local branch focuses on the information from the salient lesion regions. The weighted feature combination from both branches further boosts the classification performance of the skin lesions.

A. Skin Lesion Localization

Localizing the skin lesion in a skin lesion image is an important step for understanding the lesion structure. To learn the lesion feature from the region of interest, we design a local learning branch with a skin lesion localization module. Given an input image as $\mathbf{x}$, the feature maps after the global branch are denoted as $\{g_k\}_{k=1}^{K}$ $(H \times W \times K)$, where H, W and K are height, width, and channel of the feature maps. To extract the region that is crucial for skin lesion classification, we generate an activation map M by

$$M = \sum_{k=1}^{K} g_k / K. \tag{1}$$

Here, the activation map M indicates the importance of the region for skin lesion classification. Then, we transfer the activation map M to be a binary mask $M^{'}$ by

$$M^{'}(i,j) = \begin{cases} 1, & if \quad M(i,j) \geqslant T, \\ 0, & if \quad M(i,j) < T, \end{cases} \tag{2}$$

where T is a threshold that controls the size of the region of interest. In this paper, T is set to be $\sum_{i,j} M(i,j)/(H \times W)$, where (i,j) indicates the pixel location in the activation map M. From the binary mask $M^{'}$, we can delineate a rectangle that contains the skin lesion. With the coordinate information of that rectangle, we crop the salient lesion region $\hat{\mathbf{x}}$ from the dermoscopic image.

We input the salient lesion region $\hat{\mathbf{x}}$ into the local branch and obtain the feature maps $\{l_k\}_{k=1}^{K}$. In the local branch, our network concentrates on learning the discriminative feature from skin lesions. For dermoscopic images with noises (like hairs, rulers, and air bubbles), our local branch can effectively reduce the influence of those distractions, which further improves the feature representative ability of the skin lesions. In addition, some skin lesions in dermoscopic images are so small that is hard for the network to extract their features after a series of downsmalping operations (e.g., pooling). The proposed local branch can alleviate this issue by focusing on the lesion region for feature extraction.

B. Weighted Feature Integration

From the global and local branches of the network, we extract the feature maps $\{g_k\}_{k=1}^{K}$ and $\{l_k\}_{k=1}^{K}$ for skin lesion classification. To exploit the informative feature representation from $\{g_k\}_{k=1}^{K}$ and $\{l_k\}_{k=1}^{K}$, we combine the feature information from the global and local branches by

$$F = \mathfrak{F}_p[\{g_k\}_{k=1}^{K}, \{l_k\}_{k=1}^{k}], \tag{3}$$

where $\mathfrak{F}_p$ is a concatenation operation. Then, we learn a mutual vector μ from F by

$$\mu = \mathfrak{F}_q(F; \theta_q), \tag{4}$$

where $\mathfrak{F}_q$ is a linear mapping function. By learning how to combine the lesion feature at each channel through $\mathfrak{F}_q$, our network achieves an adaptive fusion of the information from global branch $\{g_k\}_{k=1}^{K}$ and local branch $\{l_k\}_{k=1}^{K}$. θ_q are the parameters of $\mathfrak{F}_q$. We then normalize μ as

$$\hat{\mu} = \text{sigmoid}(\mu). \tag{5}$$

To capture the representative lesion information from two branches while reduce the property degradation of the F, we weight F with the guidance of the global and local feature information $\hat{\mu}$ via residual attention, i.e.,

$$F^{'} = \mathfrak{F}_s(\hat{\mu} \times F, F), \tag{6}$$

where $\mathfrak{F}_s$ is a summation operation. A linear classifier is used to classify the input dermoscopic image as melanoma or non-melanoma. With the processes of Eqs. (3-6), the proposed network fuses the lesion information from global and local views, which enhances the discriminative clues from both branches.

III. EXPERIMENT

A. Materials

To evaluate the proposed network on skin lesion classification, we test our proposed network on the ISBI 2017 skin lesion dataset. It contains 2000 training images, 150 validation images and 600 test images. In the training set, there are 374 melanoma cases and 1626 non-melanoma cases. In the validation set, there are 30 melanoma cases and 120 non-melanoma cases. The test set includes 117 melanoma cases and 483 non-melanoma cases. The size of those dermoscopic images varies from 540×722 to 4499×6478.

B. Evaluation Criteria

We use four measurements to assess the performance of skin lesion classification, i.e., area under the receiver operating characteristic curve (AUC), accuracy (ACC), sensitivity (SEN), specificity (SPE). Metrics of ACC, SEN and SPE are defined as:

$$\begin{aligned} \text{ACC} &= (\text{TP}+\text{TN})/(\text{TP}+\text{TN}+\text{FP}+\text{FN}), \\ \text{SEN} &= \text{TP}/(\text{TP}+\text{FN}), \\ \text{SPE} &= \text{TN}/(\text{TN}+\text{FP}), \end{aligned} \tag{7}$$

where TP and TN denote the number of pixels correctly classified as melanoma and non-melanoma, respectively. FP and FN represent the number of pixels incorrectly classified as melanoma and non-melanoma, respectively.

TABLE I
MELANOMA CLASSIFICATION OF THE PROPOSED NETWORK ON ISBI 2017 DATASET (%)

Method	AUC	ACC	SEN		
			SPE=0.95	SPE=0.9	SPE=0.85
Resnet101	84.03	85.50	46.15	51.28	59.83
Densnet121	84.49	85.16	41.03	52.99	60.68
Proposed method$_1$	87.17	86.34	50.43	65.80	72.64
Proposed method$_2$	**87.84**	**87.33**	**55.56**	**68.37**	**76.06**

Proposed method$_1$ takes ResNet101 as the baseline network. Proposed method$_2$ uses DenseNet121 as the baseline network.

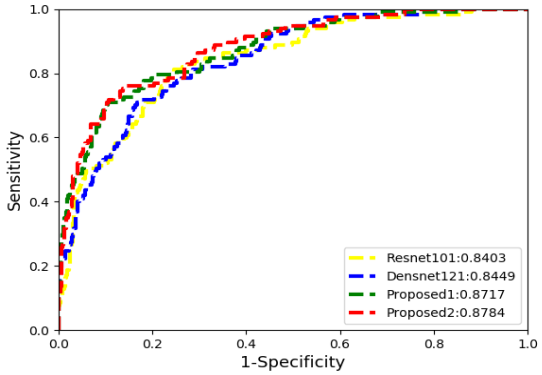


Fig. 3. ROC curves of different methods for melanoma classification.

C. Training Setting

The network can be implemented on the basis of several state-of-the-art deep networks. In this paper, we take ResNet101 [18] and DenseNet121 [19] as our baseline networks. Both baseline networks are pre-trained on ImageNet [20]. The loss function for skin lesion classification is cross-entropy loss. We use stochastic gradient descent (SGD) for network optimization. The network is trained in an end-to-end manner with 100 epochs and a batch-size of 15. For batch processing, all inputs for the global and local branches are resized to 256 × 256. Augmentation processes including random rotation, horizontal and vertical flips are performed to produce more training dermoscopic images.

D. Ablation Study

We show the melanoma classification performance based on the baseline network and the proposed network in Table 1. Compared with the baseline models on melanoma classification, our proposed networks enhance melanoma classification by about 3.1%-3.4% of AUC, which displays the effectiveness of the information combination from the global and local branches for skin lesion classification. We also list the sensitivity when the specificity is fixed as a high value in Table 1. It can be observed that the proposed method can significantly increase the sensitivity, e.g., enhancing the sensitivity by about 15% when the specificity is 0.95. In

TABLE II

PERFORMANCE COMPARISON OF MELANOMA CLASSIFICAITION ON ISBI 2017 DATASET (%)

Method	AUC	ACC	SEN	SPE
Zhang *et al*. [15]	85.90	83.70	59.00	89.60
Yang *et al*. [9]	84.20	83.00	60.70	88.40
Sultana *et al*. [12]	78.90	83.20	52.90	90.50
Al *et al*. [13]	N.A.	81.57	**75.33**	80.62
Harangi [10]	85.10	85.20	40.20	71.90
Proposed method	**87.84**	**87.33**	54.70	**95.24**

For fair comparison, only results from the same training data are recorded.

addition, our proposed strategy improves the skin lesion classification on both ResNet101 and DenseNet121, which shows the robustness of the proposed strategy on different deep networks for skin lesion classification.

Fig. 3 shows the ROC curves of the melanoma classification based on different deep networks. From Fig. 3, we can observe that the proposed strategy enhances the performance of melanoma classification consistently on two different baseline networks. Specifically, the sensitivity of the melanoma classification is improved when the specificity at a relatively high value, as shown in the red and green curves in Fig. 3.

E. Comparison with Other Methods

We also compare our proposed method with other recently published works on melanoma classification. Table 2 shows the comparison results, where we list the melanoma classification performance from the methods using the same training set without external training images. From Table 2, it can be seen that the proposed method achieves the best melanoma classification AUC, which exemplifies the validity of the proposed deep network on melanoma classification. Moreover, If we fix the specificity as 0.90, the proposed method produces the sensitivity of 0.6837, which is significantly higher than the 0.5900 ([15]), and 0.5290 ([12]). In addition, compared with other methods that use the segmentation information to localize the skin lesions (like [9], [12]), our proposed technique can localize lesion regions without the need of the annotation of lesion segmentation information. Thus, our proposed method can be directly and flexibly extended to other biomedical classification tasks.

IV. CONCLUSION

In this paper, a multi-level attentive skin lesion learning network is proposed to enhance melanoma classification. By integrating lesion localization information into the feature learning process, we boost the interpretability of the proposed network on lesion feature representation. In addition, by weighted fusing the feature information from the global and local branches, we enrich the feature representation of the skin lesion, which further improves the performance of the proposed network on melanoma classification. Experimental analysis validates the effectiveness of the proposed method on skin lesion classification.

REFERENCES

[1] X. Wang, X. Jiang, H. Ding, and J. Liu, "Bi-directional dermoscopic feature learning and multi-scale consistent decision fusion for skin lesion segmentation," *IEEE Trans. Image Process.*, vol. 29, pp. 3039–3051, 2020.

[2] X. Wang, H. Ding, and X. Jiang, "Dermoscopic image segmentation through the enhanced high-level parsing and class weighted loss," in *Proceedings of the IEEE International Conference on Image Processing*, 2019, pp. 245–249.

[3] Z. Hu, J. Tang, Z. Wang, K. Zhang, L. Zhang, and Q. Sun, "Deep learning for image-based cancer detection and diagnosis- a survey," *Pattern Recognit.*, vol. 83, pp. 134–149, 2018.

[4] N. Gessert, T. Sentker, F. Madesta, R. Schmitz, H. Kniep, I. Baltruschat, R. Werner, and A. Schlaefer, "Skin lesion classification using cnns with patch-based attention and diagnosis-guided loss weighting," *IEEE Trans. Biomed. Eng.*, 2019.

[5] Q. Kong, R. Lu, F. Yin, and S. Cui, "Privacy-preserving continuous data collection for predictive maintenance in vehicular fog-cloud," *IEEE Trans. Intell. Transp. Syst.*, 2020.

[6] Y. Yuan, M. Chao, and Y. Lo, "Automatic skin lesion segmentation using deep fully convolutional networks with jaccard distance," *IEEE Trans. Med. Imaging*, vol. 36, no. 9, pp. 1876–1886, 2017.

[7] C. Barata, M. Celebi, and J. Marques, "Explainable skin lesion diagnosis using taxonomies," *Pattern Recognit.*, p. 107413, 2020.

[8] J. Kawahara, S. Daneshvar, G. Argenziano, and G. Hamarneh, "Seven-point checklist and skin lesion classification using multitask multimodal neural nets," *IEEE J. Biomed. Health Informs.*, vol. 23, no. 2, pp. 538–546, 2018.

[9] J. Yang, F. Xie, H. Fan, Z. Jiang, and J. Liu, "Classification for dermoscopy images using convolutional neural networks based on region average pooling," *IEEE Access*, vol. 6, pp. 65130–65138, 2018.

[10] B. Harangi, "Skin lesion classification with ensembles of deep convolutional neural networks," *J. Biomed. Inform.*, vol. 86, pp. 25–32, 2018.

[11] L. Song, J. Lin, Z. Wang, and H. Wang, "An end-to-end multi-task deep learning framework for skin lesion analysis," *IEEE J. Biomed. Health Informs.*, vol. 24, no. 10, pp. 2912–2921, 2020.

[12] N. Sultana, B. Mandal, and N. Puhan, "Deep residual network with regularised fisher framework for detection of melanoma," *IET Computer Vision*, vol. 12, no. 8, pp. 1096–1104, 2018.

[13] M. Al-masni, D. Kim, and T. Kim, "Multiple skin lesions diagnostics via integrated deep convolutional networks for segmentation and classification," *Comput. Meth. Prog. Bio.*, p. 105351, 2020.

[14] L. Yu, H. Chen, Q. Dou, J. Qin, and P. Heng, "Automated melanoma recognition in dermoscopy images via very deep residual networks," *IEEE Trans. Med. Imaging*, vol. 36, no. 4, pp. 994–1004, 2017.

[15] J. Zhang, Y. Xie, Y. Xia, and C. Shen, "Attention residual learning for skin lesion classification," *IEEE Trans. Med. Imaging*, 2019.

[16] Z. Yu, X. Jiang, F. Zhou, J. Qin, D. Ni, S. Chen, B. Lei, and T. Wang, "Melanoma recognition in dermoscopy images via aggregated deep convolutional features," *IEEE Trans. Biomed. Eng.*, vol. 66, no. 4, pp. 1006–1016, 2018.

[17] Y. Gu, Z. Ge, C. Bonnington, and J. Zhou, "Progressive transfer learning and adversarial domain adaptation for cross-domain skin disease classification," *IEEE J. Biomed. Health Informs.*, 2019.

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

[19] G. Huang, Z. Liu, G. Pleiss, Van D., and K. Weinberger, "Convolutional networks with dense connectivity," *IEEE Trans. Pattern Anal. Mach. Intell.*, 2019.

[20] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, et al., "Imagenet large scale visual recognition challenge," *Int. J. Comput. Vision*, vol. 115, no. 3, pp. 211–252, 2015.