Multi-level Relationship Capture Network for Automated Skin Lesion Recognition

Zihao Liu[1,2], Ruiqin Xiong[1], and Tingting Jiang[1]

[1] NELVT, Department of Computer Science, Peking University, Beijing, China
[2] Advanced Institute of Information Technology, Peking University, Hangzhou, China
{lzh19961031,ttjiang}@pku.edu.cn

Abstract. Automated skin lesion recognition of dermoscopy images is effective for improving diagnostic performance. Current popular solutions either leverage a single image to learn better feature representations or take advantage of pairwise images for more discriminative recognition. However, they ignore modeling the relationship between important regions within the central lesion area, or mining the deeper semantic correlation between different images. In this paper, we propose a novel Multi-level Relationship Capture Network (MRCN), which focuses on relationship mining at two different levels, the region level and the image level. Specifically, a region-correlation learning module is proposed to model the relationship between different important regions in the central lesion area. Meanwhile, a cross-image learning module is designed to model the deep semantic correlation between multiple images. Besides, a lesion discerning module and a consistency regularization module are adopted to extract the feature of the lesion area and to serve as an extra consistency constraint, respectively. Comprehensive experiments are conducted on three challenging datasets, and the experimental results show that our MRCN can achieve the state-of-the-art performance compared to previous work, which demonstrates its advantages and superiority.

1 Introduction

Skin disease is one of the most common diseases in the world, which aroused public attention [15, 13]. A large number of methods have been proposed for the automated recognition of dermoscopy images since the manual inspection is subjective.

Most methods utilize a single image for the final recognition. Early approaches apply hand-crafted features to solve this problem [17, 10, 2]. Recently, many CNN-based methods are also proposed. One stream of them is mainly designed for learning better feature representations [27, 8]. Nevertheless, it is not enough to work at the feature level. For dermoscopy images, only the lesion area located in the center of the image is valuable for the diagnosis, which is called the "central lesion area", as shown in Fig. 1. Regarding this, another stream aims to take advantage of this characteristic. Some crop out the lesion area before the classification [26, 14], the others utilize attention mechanisms to focus on the lesion area [30]. However, all the above methods ignore to mine the hidden information within the central lesion area. During the diagnosis process of dermatologists, different regions within the central lesion area are examined by doctors. These regions are viewed by different importance, and the relationship of them is

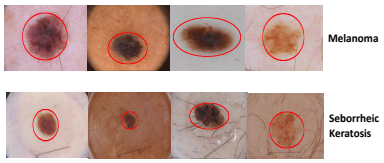

Fig. 1. Some examples of melanoma and seborrheic keratosis. The central area circled by the red circle is called the "central lesion area".

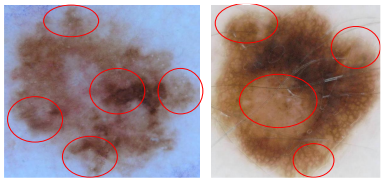

Fig. 2. The illustration of important regions. The red circle indicates the important sub-regions within the central lesion area. They are usually located in the center or on the edges and their relationship is evaluated by doctors.

evaluated for a more in-depth analysis, as illustrated in Fig. 2. Note that here "region" denotes local attended regions within the central lesion area. Thus efficiently modeling the relationship between these meaningful and important regions is important for the classification, which is called **"Region level relationship challenge"**.

Besides the region-level relationship challenge with a single image, there is another challenge at the image level. For the dermoscopy image, the visual difference within the same class could be even more notable than that between different classes, as shown in Fig. 1. How to effectively explore the semantic similarities and discriminations between different images, no matter whether they are of the same category or not, is a big challenge of this task, which is called **"Image level relationship challenge"**. To tackle this, a few recent approaches propose to utilize image pairs instead of a single image [28, 22, 20], and discriminate whether they are from the same class. However, they just simply concat the two features, ignoring to model the deeper semantic correlation between the two images for more abundant messages, which could facilitate each other. For doctors, it is commonly adopted to mine complementary information and summarize contrastive visual appearances, *e.g.*, semantic similarity and the discrimination positions with different scales and locations, as for a more effective joint judgment. Thus, there is still much room to improve the solution for the "Image level relationship challenge".

To address the above two challenges, we propose a novel Multi-level Relationship Capture Network (MRCN), which **focuses on relationship mining at two different levels, the region level and the image level.** At the region level, inspired by the attention mechanism [21] and guided by doctors' expertise, a region-correlation learning module is proposed to model the relationship between different important regions within the central lesion area. At the image level, inspired by the doctors' practice, a cross-image learning module is introduced to learn the deep semantic correlation between multiple images for complementary information. Besides, a lesion discerning module and a consistency regularization module are proposed to extract the feature of the central lesion area and serve as an extra regularization, respectively.

Experiments are conducted on three public datasets to demonstrate the effectiveness of our MRCN. We achieve state-of-the-art performance on all of them. To sum up, the main contributions are: (1) To our best knowledge, this is the first paper to mine the relationship both at the image level and the region level for this task. The correlation between different important regions of the lesion area is modeled, and the deep seman-

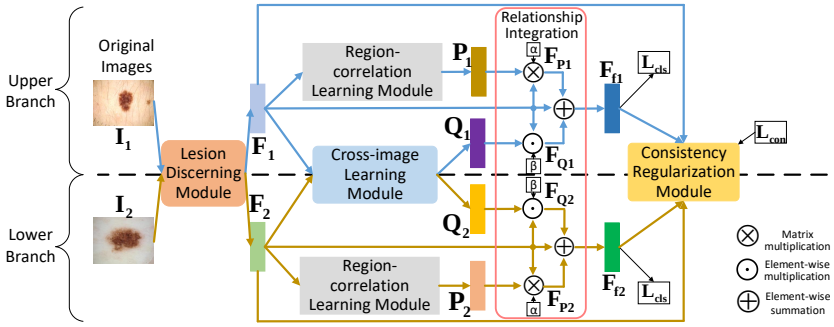

Fig. 3. An overview of the MRCN. There are two branches, the upper and the lower branch.

tic correlation between multiple images is learned to facilitate each other. (2) A new architecture of MRCN is proposed, including two newly designed modules: region-correlation learning module and cross-image learning module, which are deeply in line with the intuition of doctors and integrates their expertise. (3) Our MRCN achieves state-of-the-art performance on three public datasets.

2 Methodology

In this section, we elaborate on the whole architecture of MRCN, which is illustrated in Fig. 3. Given an image pair I_1 and I_2, they are first processed by the lesion discerning module, generating the features corresponding to the central lesion areas for each branch, denoted as F_1 and F_2. The following region-correlation learning module is equipped by each branch. It takes F of each image as input and outputs an attention feature P for each image, denoted as P_1 and P_2. In parallel, a cross-image learning module is proposed, which synergically utilizes F_1 and F_2 as input. An attention feature Q is generated for each image, denoted as Q_1 and Q_2. After that, for each branch, P and Q are aggregated with F for relationship integration, obtaining the final feature F_f, which is used for the final recognition for each image. Finally, serving as an extra regularization, the consistency regularization module takes F_1, F_2, F_{f1}, and F_{f2} as input, and evaluate the consistency, *i.e.*, whether they belong to the same image.

Lesion Discerning Module. This module is introduced to extract the feature of the "central lesion area", as shown in Fig. 4. It contains two parts: the lesion attention part, to crop the central lesion area; and the feature extraction part, to extract the feature.

The lesion attention part includes four conv blocks and four deconv layers. Each conv block contains three conv layers, with a batch normalization layer and ReLU layer after each conv layer. It will output a lesion attention map, which is the binary segmentation result for the image. After that, the smallest rectangle which includes the lesion area is taken from the attention map to crop out the original image. The cropped lesion patch, as the input of the feature extraction part, is fed into a CNN backbone to extract the original feature $F \in \mathbb{R}^{H \times W \times C}$, which is the output of this module.

Region-correlation Learning Module This module is designed to mine the region level relationship, as shown in Fig. 5. Previous works have demonstrated that in CNNs,

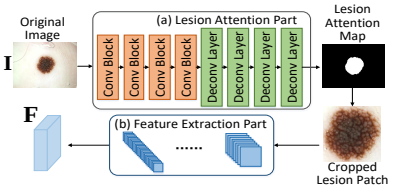

Fig. 4. Lesion discerning module.

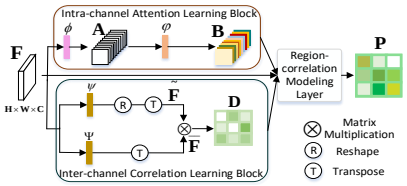

Fig. 5. Region-correlation learning module.

different feature channels correspond to different locations and regions of an image [25, 16]. Thus, we design two channel-wise based attention blocks. The intra-channel attention learning block evaluates the importance of each channel itself, and the inter-channel correlation block models the correlation between channels. After that, the information of these two blocks is integrated by a region-correlation modeling layer.

For intra-channel attention learning block, firstly, an average pooling function ϕ is applied on F, which generates $A \in \mathbb{R}^{C\times 1}$. Then a learning function φ will be applied to A to further study the importance of each channel and generate $B \in \mathbb{R}^{C\times 1}$. In B, each element b_k represents the importance of k^{th} channel.

$$A = \phi(F) = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} F_{ij}, \quad B = \varphi(A) \tag{1}$$

On the other hand, for inter-channel correlation block,F will go through two learning functions ψ and Ψ respectively, then a matrix multiplication between the two results is conducted to generate an attention map $D \in \mathbb{R}^{C\times C}$:

$$D = \psi(F) * \Psi(F) = \widetilde{F} * \overline{F} \tag{2}$$

where "*" means matrix multiplication. In D, each element d_{ij} measures the j^{th} channel's impact on the i^{th} channel.

Next, the region-correlation modeling layer is introduced to merge the messages of B and D and outputs $P \in \mathbb{R}^{C\times C}$. Each element p_{ij} of P is calculated by Eqn. (3):

$$p_{ij} = d_{ij}(b_i b_j) \tag{3}$$

In this way, the importance of each channel itself will be integrated into the relationship between channels. It is worth noting that [21] simply models the spatial and channel relationships by pooling. But this module focuses on the correlation between important regions by integrating the importance of each channel and the relationship between them. The motivation, perspective and architecture are different.

Cross-image Learning Module This module is designed to model the image level relationship, as illustrated in Fig. 6. Firstly, F_1 and F_2 is processed by two learning functions T_1 and T_2 respectively and obtain $U \in \mathbb{R}^{N\times C}$ and $V \in \mathbb{R}^{N\times C}$, where $N(N = HW)$ is the number of spatial positions. Then, the spatial correlation modeling layer further measures the contextual semantic relevance. Lastly, the complementary learner encodes the learned correlation and generates the final attention maps.

The spatial modeling layer takes U and V as input, measures the correlation by cosine distance. It obtains two spatial correlation maps $S^{2\to1}$, $S^{1\to2} \in \mathbb{R}^{N\times N}$:

$$S_{ij}^{2\to1} = (\frac{u_i}{||u_i||_2})(\frac{v_j}{||v_j||_2})^T, \quad S_{ij}^{1\to2} = (\frac{v_i}{||v_i||_2})(\frac{u_j}{||u_j||_2})^T, \quad i,j = 1,...,N \tag{4}$$

in which u_i denotes the i^{th} row in U, similarly with v_i. Each $S_{ij}^{2\to1}$ is an affinity score reflects the message from j^{th} element in V to the i^{th} element in U. Similarly with $S^{1\to2}$. Therefore, for one image, the semantic relevant elements in the other image are highlighted, results in higher values for the corresponding elements in $S^{2\to1}$ and $S^{1\to2}$.

The following complementary learner consists of two parts. The first part takes $S^{2\to1}$ and $S^{1\to2}$ as input, further learns the relevance between different elements by a learning function ϑ, encodes the message and generates attention map $Q_1 \in \mathbb{R}^{N\times1}$ and $Q_2 \in \mathbb{R}^{N\times1}$, which is the output of this module.

$$Q_1(i) = \sum_j^N \vartheta(S_{ij}^{2\to1}), \qquad Q_2(i) = \sum_j^N \vartheta(S_{ij}^{1\to2}) \tag{5}$$

For Q_1, the i^{th} element represents the integratation of the semantic messages from all the elements in F_2 to the i^{th} element in F_1. Similarly with Q_2. The learning functions, which could be seen as the combination of conv functions and activation functions, are illustrated specifically in Sec. 3.2. Therefore, those elements with a higher response, indicating more correlation with the other image, will correspond to a higher final score, emphasizing the complementary information.

Relationship Intergration For each branch, P and Q will be integrated to F, as illustrated in Fig. 3. A matrix multiplication is applyed between P and F, meanwhile Q will be fused to original feature F by appling an element-wise multiplication. Then the results will be multipled by a learnable weighting factor α and β respectively, then conduct a matrix summation to F, to generate the final feature F_f for each branch:

$$F_f = F + \alpha \cdot (P * F) + \beta \cdot (Q \cdot F) \tag{6}$$

where "$\cdot$" is element-wise multiplication and "*" denotes matrix multiplication.

F_f will pass through two fully connected layers to obtain the classification probabilistic prediction, which is supervised by normal cross-entropy loss $\mathcal{L}_{\text{cls}}$.

Consistency Regularization Module An extra regularization is needed to impose extra constraints in addition to the effect of the Cross-image learning module and

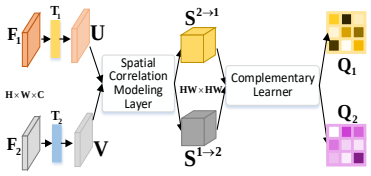

Fig. 6. Cross-image learning module.

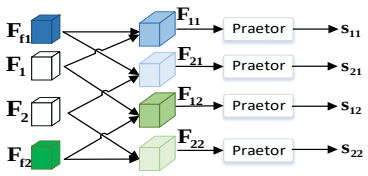

Fig. 7. Consistency regularization module.

Region-correlation learning module for better feature learning. Specifically, the constraints are to ensure that after integrating messages from other images, the network can still correctly discriminate which pair of F and F_f is from the same image and which is from different images, illustrated as Fig. 7.

Taking F_1, F_2, F_{f1}, F_{f2} as input, the concatenation of F_1 and F_{f1} (F_{11}); F_2 and F_{f1} (F_{21}); F_1 and F_{f2} (F_{12}); F_2 and F_{f2} (F_{22}), are fed into an adaptive "praetor" respectively, and output four consistency scores. The praetor consists of two convolutional layers and two fully connected layers. The consistency score s_{11}, which corresponds to F_{11} is optimized to as close as possible to 1 since F_1 and F_{f1} are from the same image. The same with s_{22}. On the other hand, the consistency scores s_{12} and s_{21} are expected to be close to 0. These four scores will be supervised by binary cross-entropy loss:

$$\mathcal{L}_{\text{con}} = -\frac{1}{N_p}\sum_{z=1}^{N_p}\sum_{i=1}^{2}\sum_{j=1}^{2}(y_{ij}^{z}\log(s_{ij}^{z}) + (1-y_{ij}^{z})\log(1-s_{ij}^{z})) \tag{7}$$

where y_{ij}^{z} is the consistency label of z^{th} image pair, N_p is the total number of image pair. For each image pair, y_{11}^{z} and y_{22}^{z} are 1, and y_{12}^{z} and y_{21}^{z} are 0.

The final loss is computed as $\mathcal{L} = \mathcal{L}_{\text{cls}} + \gamma\mathcal{L}_{\text{con}}$, where γ is a hyper-parameter.

3 Experiments

3.1 Datasets

We employ three benchmark datasets for experiments: the ISIC 2016 challenge dataset [7] consisting of 1279 images from 2 categories, the ISIC 2017 challenge dataset [3] including 2750 images from 3 categories and the ISIC 2019 challenge dataset [18, 4] including 33569 images from 9 classes. We use the official training set, validation set and test set for evaluation. Note that ISIC 2016 and ISIC 2017 are two ended challenges, ISIC 2019 is an ongoing challenge, the results are obtained by submitting the predictions to the platform [19], which will be published on the leaderboard.

3.2 Implementation Details

ResNet50 is chosen as the backbone for the feature extraction part. For each image, its pair image is randomly chosen, with each resized to 448 $\times$ 448. ψ and Ψ are 1 $\times$ 1 convolution layers, T_1 and T_2 are conducted by 3 $\times$ 3 convolution layers. φ is the combination of a conv layer and a ReLU layer, and ϑ is the combination of two conv layers, with a ReLU layer between them. The learning rate is initialized to 0.001 and annealed by 0.5 every 10 epochs. The batch size is set to 40 on four NVIDIA GTX 2080Ti GPUs. γ is set to 0.05. As for evaluation metrics, we utilize Area Under Receiver operation Curve (AUC), Average Precision (AP), Accuracy (ACC), Sensitivity (SE) and Specificity (SP). **Training phase**: We follow the tradition in the other methods [6] and use the data with segmentation maps to train the lesion attention part of the lesion discerning module separately first at the training phase. After that, taking an image pair as inputs, the lesion discerning module outputs two corresponding features.

These two features are the input of later architecture. **Testing phase**: The cross-image learning module and consistency regularization module will be removed during inference. Taking a single test image for the input, only the upper branch is used to obtain the final result.

3.3 Ablation Study

To investigate the impact of all the components in the network, we apply the ablation study on ISIC 2016 dataset. The performance is shown in Table 1. We denote "LD" as lesion discerning module, "RC" as region-correlation learning module, "CL" as cross-image learning module and "CR" as consistency regularization module. The pre-trained ResNet50 is used as the baseline model, denoted as "Baseline", which obtains an AP of 0.698, ACC of 0.843 and AUC of 0.814. The third, fourth and fiveth rows are respectively the results of adding "LD", "RC" and "CL" to the baseline. The consistency regularization module will only work when there is at least one of the region-correlation learning module and the cross-image learning module. The AP value is improved by 3.2%, 3.5% and 3.3% respectively. The experimental result proves that individually adopting the three modules can benefit the model. When two modules are combined with baseline, illustrated as "LD+RC", "LD+CL", "RC+CR" and "CL+CR", the results are better. This demonstrates that combining two modules performs better than only combining one module with baseline. On this basis, when three modules are adopted, illustrated as "LD+RC+CL" , "LD+RC+CR", "LD+CL+CR", and "RC+CL+CR", the performance further improves, proving that compared to combining two modules, adopting three modules gain a further improved performance. Finally, when the full model is adopted, represented as "MRCN (full)", the performance is the best, improving over the baseline by 10.9%, 5.5%, 7.6% in AP, ACC and AUC. Besides, the improvement of CR is smaller compared to CL and RC. For example, "LD+RC+CR" is worse than "LD+RC+CL"; "LD+CL+CR" is worse than "LD+RC+CL". This demonstrates the effectiveness of CL, RC themselves, and that CR only serves as an extra regularization.

3.4 Comparison with Other Methods

ISIC2016. To follow the tradition of other methods, we compare the performance of AP, ACC and AUC with five recent methods and top-five ranking methods on the challenge leaderboard. The results are shown in Table 2. This challenge is ranked based only on AP. Also, we do not use any extra data. Our method achieves the best performance in all three metrics, in which the AP is 0.807, significantly surpass the second place by 6.7%. Our ACC and AUC also exceed DCNN-FV by 1.1% and 2.7%.

ISIC2017. We compare our method with six recent methods and top-five ranking methods on the challenge leaderboard. The results are shown in Table 3. Note that the Average AUC of the two sub-tasks is the ranking metric of this challenge. The AUC are 0.947 and 0.988 respectively for the two sub-tasks, which improve the second place by 2.7% and 0.7%, and the average AUC improves by 1.8%. In addition, the results of AP, ACC in two sub-tasks, and SE in sub-task1, are also best, comparing to other

Table 1. Ablation Study on ISIC 2016 dataset.

Method	AP	ACC	AUC
Baseline	0.698	0.843	0.824
LD	0.730	0.848	0.839
RC	0.733	0.852	0.834
CL	0.731	0.861	0.846
LD+RC	0.757	0.871	0.862
LD+CL	0.760	0.872	0.863
RC+CR	0.749	0.873	0.868
CL+CR	0.752	0.873	0.868
LD+RC+CL	0.784	0.880	0.891
LD+RC+CR	0.776	0.882	0.879
LD+CL+CR	0.781	0.881	0.883
RC+CL+CR	0.779	0.883	0.886
MRCN (full)	**0.807**	**0.898**	**0.900**

Table 2. Results of our method, five recent methods and top five ranking methods on ISIC 2016. "AP" is the only ranking metric.

Method	AP^*	ACC	AUC
Our MRCN	**0.807**	**0.898**	**0.900**
CIN [9]	0.740	0.887	0.873
L-CNN [20]	0.724	0.876	0.854
AttnMel-CNN [24]	0.693	-	0.852
DCNN-FV [27]	0.685	0.868	0.852
SDL [28]	0.664	0.858	0.818
CUMED [26]	0.637	0.855	0.804
GTDL [7]	0.619	0.813	0.802
Result2 [7]	0.615	0.844	0.808
USYD [7]	0.580	0.686	0.793
Mufic-IT [7]	0.534	0.760	0.685

Table 3. Results of our method, six recent methods and top five ranking methods on ISIC 2017 Dataset. Note that "Average AUC" is the only ranking metric, which is highlighted by "*".

Methods	External data	Melanoma Classification					Seborrheic Keratosis					*Average*
		AUC^*	AP	ACC	SE	SP	AUC^*	AP	ACC	SE	SP	AUC^*
Our MRCN	0	**0.947**	**0.864**	**0.906**	**0.796**	0.921	**0.988**	**0.917**	**0.949**	0.918	0.946	**0.968**
CIN [9]	0	0.920	0.814	0.894	0.645	0.948	0.981	0.902	0.943	0.829	0.965	0.951
MBDCNN [23]	1320	0.903	-	0.878	0.727	0.915	0.973	-	0.93	0.844	0.945	0.938
ARL-CNN [30]	1320	0.875	-	0.850	0.658	0.896	0.958	-	0.868	0.878	0.867	0.917
SSAC [22]	1320	0.873	-	0.835	0.556	0.903	0.959	-	0.912	0.889	0.916	0.916
SDL [29]	1320	0.868	0.689	0.872	-	-	0.955	0.818	0.917	-	-	0.912
RENI [11]	1444	0.868	0.710	0.828	0.735	0.851	0.953	0.786	0.803	**0.978**	0.773	0.911
gpm-LSSSD [5]	900	0.856	0.747	0.823	0.103	**0.998**	0.963	0.839	0.875	0.178	**0.998**	0.910
Alea-Jacta-Est [12]	7544	0.874	0.715	0.872	0.547	0.950	0.943	0.790	0.895	0.356	0.990	0.908
EResNet [1]	1600	0.870	0.732	0.858	0.427	0.963	0.921	0.770	0.918	0.589	0.976	0.896

approaches. To sum up, without using any extra data, our method achieves the best performance on the ranking metric and most of the other metrics.

ISIC2019. This challenge dataset is ranked based only on the balanced multi-class accuracy (BMCA), which is the average recall score. We compare with the methods on the challenge leaderboard. The results are shown on the platform [19]. Our method obtains the highest BMCA with 0.635, noticeably improved the second place by 1.2%. Besides, we yield the best result on SE, NPV and the second in AUC.

3.5 Visualization Results

To better understand how region-correlation learning module and cross-image learning module work, we visualize the attention maps of P and Q, and show three examples in Fig. 8. As shown, the attention maps P successfully pay attention to the regions on the edges, as well as those located in the center, which are both crucial for the diagnosis. Besides, comparing the attention maps of P and Q for the same image, the

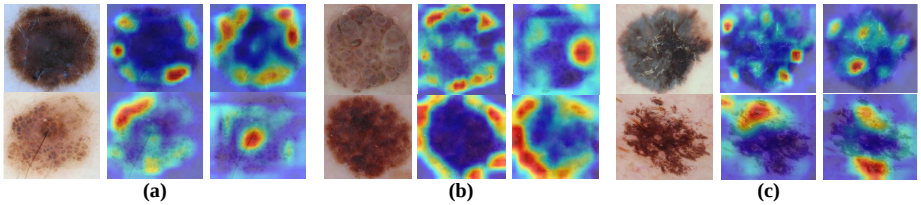

Fig. 8. Visualization results of attention map P and Q. There are three examples, each example contains three columns: the original image pair, the corresponding activation of P and the corresponding activation of Q.

attention map of Q can identify some central and edge regions which have not been highlighted by P. This result suggests that the cross-image learning module can learn useful supplementary messages between image pairs.

4 Conclusion and Future work

In this paper, we propose a novel Multi-level Relationship Capture Network (MRCN), which focuses on relationship mining from two levels, the region and the image level. Specifically, it contains four modules, a lesion discerning module, a region-relation learning module, a cross-image learning module and a consistency regularization module. The proposed method achieves state-of-the-art performance on three benchmark datasets. In future works, we will give more qualitative and quantitative results, including discussions of each module in our method, and of the top-ranking methods.

Acknowledgement. This work was partially supported by the Natural Science Foundation of China under contracts 62088102 and 62072009. We also acknowledge the Clinical Medicine Plus X-Young Scholars Project, and High-Performance Computing Platform of Peking University for providing computational resources.

References

1. Bi, L., Kim, J., Ahn, E., Feng, D.: Automated skin lesion analysis using large-scale dermoscopy images and deep residual networks. arXiv preprint arXiv:1703.04197 (2017)
2. Catarina, B., M Emre, C., Jorge S, M.: Improving dermoscopy image classification using color constancy. IEEE Journal of Biomedical and Health Informatics **19**(3), 1146–1152 (2014)
3. Codella, N.C., Gutman, D., Celebi, M.E., Helba, B., Marchetti, M.A., Dusza, S.W., Kalloo, A., Liopyris, K., Mishra, N., Kittler, H., et al.: Skin lesion analysis toward melanoma detection: A challenge at the 2017 international symposium on biomedical imaging (ISBI), hosted by the international skin imaging collaboration (ISIC). In: International Symposium on Biomedical Imaging. pp. 168–172. IEEE (2018)
4. Combalia, M., Codella, N.C., Rotemberg, V., et al.: BCN20000: Dermoscopic lesions in the wild. arXiv preprint arXiv:1908.02288 (2019)
5. Díaz, I.G.: Incorporating the knowledge of dermatologists to convolutional neural networks for the diagnosis of skin lesions. arXiv preprint arXiv:1703.01976 (2017)

6. Gutman, D., Codella, N.C., Celebi, E., Helba, B., Marchetti, M., Mishra, N., Halpern, A.: Https://challenge.isic-archive.com/landing/2016/41 (2016)
7. Gutman, D., Codella, N.C., Celebi, E., Helba, B., Marchetti, M., Mishra, N., Halpern, A.: Skin lesion analysis toward melanoma detection: A challenge at the international symposium on biomedical imaging (ISBI) 2016, hosted by the international skin imaging collaboration (ISIC). arXiv preprint arXiv:1605.01397 (2016)
8. Liang, R., Wu, Q., Yang, X.: Multi-pooling attention learning for melanoma recognition. In: 2019 Digital Image Computing: Techniques and Applications (DICTA). pp. 1–6. IEEE (2019)
9. Liu, Z., Xiong, R., Jiang, T.: Clinical-inspired network for skin lesion recognition. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 340–350. Springer (2020)
10. Margarida, R., Catarina, B., Jorge S, M., Jorge, R.: A system for the detection of melanomas in dermoscopy images using shape and symmetry features. Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization **5**(2), 127–137 (2017)
11. Matsunaga, K., Hamada, A., Minagawa, A., Koga, H.: Image classification of melanoma, nevus and seborrheic keratosis by deep neural network ensemble. arXiv preprint arXiv:1703.03108 (2017)
12. Menegola, A., Tavares, J., Fornaciali, M., Li, L.T., Avila, S., Valle, E.: RECOD titans at ISIC challenge 2017. arXiv preprint arXiv:1703.04819 (2017)
13. Rebecca L, S., Kimberly D, M., Ahmedin, J.: Cancer statistics, 2016. JAMA Dermatology **66**(1), 7–30 (2016)
14. ur Rehman, M., Khan, S.H., Rizvi, S.D., Abbas, Z., Zafar, A.: Classification of skin lesion by interference of segmentation and convolotion neural network. In: 2018 2nd International Conference on Engineering Innovation (ICEI). pp. 81–85. IEEE (2018)
15. Siegel, R.L., Miller, K.D., Jemal, A.: Cancer statistics, 2015. CA: A Cancer Journal for Clinicians **65**(1), 5–29 (2015)
16. Simon, M., Rodner, E.: Neural activation constellations: Unsupervised part model discovery with convolutional networks. In: Proceedings of the IEEE International Conference on Computer Vision. pp. 1143–1151 (2015)
17. Tatiana, T., Elisabetta La, T., Barbara, C.: Melanoma recognition using representative and discriminative kernel classifiers. In: International Workshop on Computer Vision Approaches to Medical Image Analysis. pp. 1–12. Springer (2006)
18. Tschandl, P., Rosendahl, C., Kittler, H., H: The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions. Scientific Data **5**, 180161 (2018)
19. Tschandl, P., Rosendahl, C., Kittler, H., H: ISIC 2019 live leaderboard. https://challenge.isic-archive.com/leaderboards/live (2018)
20. Wei, L., Ding, K., Hu, H., H: Automatic skin cancer detection in dermoscopy images based on ensemble lightweight deep learning network. IEEE Access (2020)
21. Woo, S., Park, J., Lee, J.Y., Kweon, I.S.: Cbam: Convolutional block attention module. In: Proceedings of the European Conference on Computer Vision. pp. 3–19 (2018)
22. Xie, Y., Zhang, J., Xia, Y.: Semi-supervised adversarial model for benign–malignant lung nodule classification on chest ct. Medical Image Analysis **57**, 237–248 (2019)
23. Xie, Y., Zhang, J., Xia, Y., Shen, C.: A mutual bootstrapping model for automated skin lesion segmentation and classification. IEEE Transactions on Medical Imaging (2020)
24. Yan, Y., Kawahara, J., Hamarneh, G., H: Melanoma recognition via visual attention. In: International Conference on Information Processing in Medical Imaging. pp. 793–804. Springer (2019)
25. Yosinski, J., Clune, J., Nguyen, A., Fuchs, T., Lipson, H.: Understanding neural networks through deep visualization. arXiv preprint arXiv:1506.06579 (2015)

26. Yu, L., Chen, H., Dou, Q., Qin, J., Heng, P.A.: Automated melanoma recognition in dermoscopy images via very deep residual networks. IEEE Transactions on Medical Imaging **36**(4), 994–1004 (2017)
27. Yu, Z., Jiang, X., Zhou, F., Qin, J., Ni, D., Chen, S., Lei, B., Wang, T.: Melanoma recognition in dermoscopy images via aggregated deep convolutional features. IEEE Transactions on Biomedical Engineering **66**(4), 1006–1016 (2019)
28. Zhang, J., Xie, Y., Wu, Q., Xia, Y.: Skin lesion classification in dermoscopy images using synergic deep learning. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 12–20. Springer (2018)
29. Zhang, J., Xie, Y., Wu, Q., Xia, Y.: Medical image classification using synergic deep learning. Medical Image Analysis **54**, 10–19 (2019)
30. Zhang, J., Xie, Y., Xia, Y., Shen, C.: Attention residual learning for skin lesion classification. IEEE Transactions on Medical Imaging **38**(9), 2092–2103 (2019)