



Skin Lesion Classification in Dermoscopy Images Using Synergic Deep Learning

Jianpeng Zhang¹, Yutong Xie¹, Qi Wu², and Yong Xia¹(✉)

¹ School of Computer Science and Engineering, Northwestern Polytechnical University, Xi'an 710072, People's Republic of China
yxia@nwpu.edu.cn

² School of Computer Science, University of Adelaide, Adelaide, SA 5005, Australia

Abstract. Automated skin lesion classification in the dermoscopy images is an essential way to improve diagnostic performance and reduce melanoma deaths. Although deep learning has shown proven advantages over traditional methods, which rely on handcrafted features, in image classification, it remains challenging to classify skin lesions due to the significant intra-class variation and inter-class similarity. In this paper, we propose a synergic deep learning (SDL) model to address this issue, which not only uses dual deep convolutional neural networks (DCNNs) but also enables them to mutually learn from each other. Specifically, we concatenate the image representation learned by both DCNNs as the input of a synergic network, which has a fully connected structure and predicts whether the pair of input images belong to the same class. We train the SDL model in the end-to-end manner under the supervision of the classification error in each DCNN and the synergic error. We evaluated our SDL model on the ISIC 2016 Skin Lesion Classification dataset and achieved the state-of-the-art performance.

1 Introduction

Skin cancer is one of the most common form of cancers in the United States and many other countries, with 5 million cases occurring annually [1]. Dermoscopy, a recent technique of visual inspection that both magnifies the skin and eliminates surface reflection, is one of the essential means to improve diagnostic performance and reduce melanoma deaths [2]. Classifying the melanoma in dermoscopy images is a significant and challenging task in the computer-aided diagnosis.

Recently, deep learning has led to tremendous success in skin lesion classification [3–5]. Ge et al. [3] demonstrated the effectiveness of cross-modality learning of deep convolutional neural networks (DCNNs) by jointly using the dermoscopy and clinical skin images. Yu et al. [4] proposed to leverage very deep DCNNs for automated melanoma recognition in dermoscopy images in two steps – segmentation and classification. Esteva et al. [5] trained a DCNN using 129,450 clinical images for the diagnose of the malignant carcinomas and malignant melanomas and achieved the performance that matches the performance of

21 board-certified dermatologists. Despite the achievements, this task remains challenging due to two reasons. First, deep neural network models may overfit the training data, as there is usually a relative small dermoscopy image dataset and this relates to the work required in acquiring the image data and then in image annotation [6]. Second, the intra-class variation and inter-class similarity pose even greater challenges to the differentiation of malignant skin lesions from benign ones [4]. As shown in Fig. 1, there is a big visual difference between the benign skin lesions (a) and (b) and between malignant lesions (c) and (d). Nevertheless, the benign skin lesions (a) and (b) are similar to the malignant lesions (c) and (d), respectively, in both shape and color.

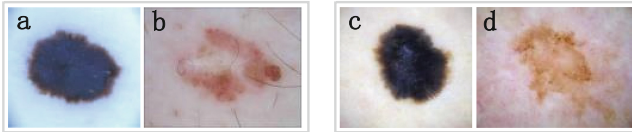


Fig. 1. Examples show the intra-class variation and inter-class similarity in skin lesion classification: (a, b) benign skin lesions, and (c, d) malignant skin lesions

To address the first issue, pre-trained deep models have been adopted, since it has been widely recognized that the image representation ability learned from large-scale datasets, such as ImageNet, can be efficiently transferred to generic visual recognition tasks, where the training data is limited [7–9]. However, it is still difficult to well address the second issue, despite some attempts reported in [4, 10]. Since hard cases (see Fig. 1) may provide more discriminatory information than easy ones [11], we, inspired by the biometrical authentication, suggest using dual DCNNs to learn from pairs of images such that the misclassification of a hard case leads to a synergic error, which can then be used to further supervise the training of both DCNN components.

Specifically, we propose a synergic deep learning (SDL) model for skin lesion classification in dermoscopy images, which consists of dual pre-trained DCNNs and a synergic network. The main uniqueness of this model includes: (1) the dual DCNNs learn the image representation simultaneously from pairs of images, including two similar images in different categories and two dissimilar images in the same category; (2) the synergic network, which has a fully connected structure, takes the concatenation the image representation learned by both DCNNs as an input and predicts whether the pair of images belong to the same class; (3) the end-to-end training of the model is supervised by both the classification error in each DCNN and the synergic error; and (4) the synergic error, which occurs when at least one DCNN misclassify an image, enables dual DCNNs to mutually facilitate each other during the learning. We evaluated our SDL model on the ISIC 2016 Skin Lesion Classification dataset [2] and achieved an accuracy of 85.75% and an average precision of 0.664, which is the current state-of-the-art.

2 Datasets

The ISIC 2016 Skin Lesion Classification dataset [2], released by the International Skin Imaging Collaboration (ISIC), is made up of 900 training and 379 test images which are screened for both privacy and quality assurance. Lesions in dermoscopic images are all paired with a gold standard (definitive) malignancy diagnosis, i.e. benign or malignant. The training set is comprised of 727 benign lesion images and 173 malignant lesion images, and the test set consists of 304 benign and 75 malignant ones.

3 Method

The proposed SDL model (see Fig. 2) consists of three modules: an input layer, dual DCNN components (DCNN-A/B) and a synergic network. The input layer takes a pair of images as input. Each DCNN component serves to learn independently the images representation under the supervision of class labels. The synergic network verifies whether the input image pair belongs to the same category or not and gives the corrective feedback if a synergic error occurs. We now delve into each of the three modules.

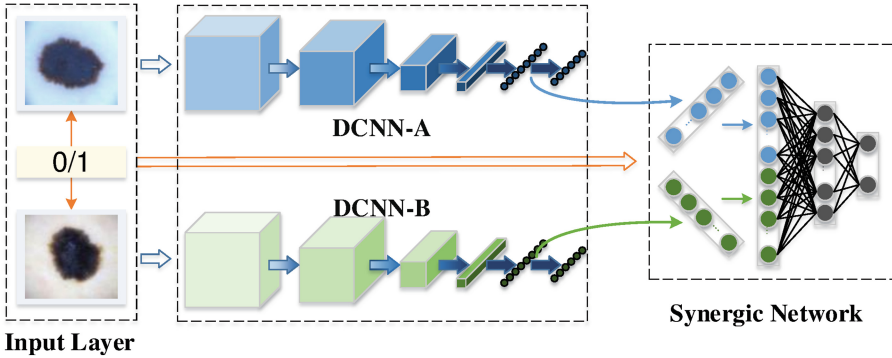


Fig. 2. Architecture of the proposed SDL model which has an input layer, dual DCNN components (DCNN-A/B) and a synergic network.

3.1 Input Layer

Different from traditional DCNNs, the proposed SDL model accepts a pair of images as an input, which are randomly selected from the training set. Each image, together with its class label, is put into a DCNN component, and each pair of images has a corresponding synergic label that is fed into the synergic network. To unify the image size, we resized each image to $224 \times 224 \times 3$ using the bicubic interpolation.

3.2 Dual DCNN Components

Although a DCNN with any structures can be embedded in the SDL model as a DCNN component, we chose a pre-trained residual network with 50 learnable layer (ResNet-50) [12] for both DCNN components, due to the trade-off between the image classification performance and the number of parameters. To adapt ResNet-50 to our problem, we replaced the original classification layer with a fully connected layer of 1024 neurons, a fully connected layer of K (the number of classes) neurons and a softmax layer, and initialized the parameters of these layers by sampling a uniform distribution $U(-0.05, 0.05)$. Then, we used an image sequence $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_M\}$ and a corresponding class label sequence $\mathbf{Y} = \{y_1, y_2, \dots, y_M\}$ to fine-tune each DCNN component, aiming to find a set of parameters θ that minimizes the following cross-entropy loss

$$l(\theta) = -\frac{1}{M} \left[\sum_{i=1}^M \sum_{j=1}^K 1\{y_i = j\} \log p(y_i = j | \mathbf{x}_i; \theta) \right] \quad (1)$$

We adopted the mini-batch stochastic gradient descent (mini-batch SGD) algorithm with a batch size of 32 as the optimizer. The obtained parameter sets for DCNN-A and DCNN-B are denoted by $\theta^{(A)}$ and $\theta^{(B)}$, respectively, which are not shared between two DCNNs during the optimization.

3.3 Synergic Network

The synergic network consists of an embedding layer, a fully connected learning layer and an output layer (see Fig. 2). Let a pair of images $(\mathbf{x}_i, \mathbf{x}_j)$ be an input of the dual DCNNs. We defined the output of the penultimate fully connected layer in DCNN-A and DCNN-B during the forward computing as the deep feature learned on image $\mathbf{x}_i, \mathbf{x}_j$, formally shown as follows

$$\begin{aligned} \mathbf{f}_i &= \mathcal{F}(\theta^{(A)}, \mathbf{x}_i) \\ \mathbf{f}_j &= \mathcal{F}(\theta^{(B)}, \mathbf{x}_j) \end{aligned} \quad (2)$$

Then, we concatenated the deep features learned on both images as an input of the synergic network, denoted by $\mathbf{f}_{i \circ j}$, and defined the expected output, i.e. the synergic label of the image pair, as

$$y_S(\mathbf{x}_i, \mathbf{x}_j) = \begin{cases} 1 & \text{if } y_i = y_j \\ 0 & \text{if } y_i \neq y_j \end{cases}. \quad (3)$$

To avoid the unbalance data problem, we set the percentage of intra-class image pairs is about 45%-55% in each batch. It is convenient to monitor the synergic signal by adding another sigmoid layer and using the following binary cross entropy loss

$$l^{(S)}(\theta^{(S)}) = y_S \log p(y_S = 0 | \mathbf{f}_{i \circ j}; \theta^{(S)}) + (1 - y_S) \log p(y_S = 1 | \mathbf{f}_{i \circ j}; \theta^{(S)}) \quad (4)$$

where $\theta^{(S)}$ is the parameters of the synergic network. If one DCNN makes a correct decision, the mistake made by the other DCNN leads to a synergic error that serves as an extra force to learn the discriminative representation. The synergic network enables dual DCNNs to mutually facilitate each other during the training process.

3.4 Training and Testing

We applied data augmentation (DA), including the random rotation and horizontal and vertical flips, to the training data, aiming to enlarge the dataset and hence alleviate the overfitting of our model. We denoted two image batches as $\mathbf{X}_A = \{\mathbf{x}_{A1}, \mathbf{x}_{A2}, \dots, \mathbf{x}_{AM}\}$, $\mathbf{X}_B = \{\mathbf{x}_{B1}, \mathbf{x}_{B2}, \dots, \mathbf{x}_{BM}\}$, corresponding classification labels as \mathbf{Y}_A , \mathbf{Y}_B , and the synergic label as \mathbf{Y}_S . After the forward computation of both DCNNs, we have two sets of deep features \mathbf{F}_A and \mathbf{F}_B . Then, we concatenated the corresponding pair of deep feature maps, and obtained $\mathbf{F}_{A \circ B} = \{\mathbf{f}_{A1 \circ B1}, \mathbf{f}_{A2 \circ B2}, \dots, \mathbf{f}_{AM \circ BM}\}$, which was used as the input of the synergic network. Next, we computed the two classification losses $l^{(A)}(\theta^{(A)})$, $l^{(B)}(\theta^{(B)})$ and synergic loss $l^{(S)}(\theta^{(S)})$ which are all cross-entropy loss. The parameters of each DCNN component and the synergic network are updated as

$$\begin{aligned}\theta^{(A)}(t+1) &= \theta^{(A)}(t) - \eta(t) \cdot \Delta^{(A)} \\ \theta^{(B)}(t+1) &= \theta^{(B)}(t) - \eta(t) \cdot \Delta^{(B)} \\ \theta^{(S)}(t+1) &= \theta^{(S)}(t) - \eta(t) \cdot \Delta^{(S)}\end{aligned}\tag{5}$$

where $\Delta^{(A)} = \frac{\partial l^{(A)}(\theta^{(A)})}{\partial \theta^{(A)}} + \lambda \Delta^{(S)}$, $\Delta^{(B)} = \frac{\partial l^{(B)}(\theta^{(B)})}{\partial \theta^{(B)}} + \lambda \Delta^{(S)}$, $\Delta^{(S)} = \frac{\partial l^{(S)}(\theta^{(S)})}{\partial \theta^{(S)}}$, λ represents the trade-off between subversion of classification error and synergic error, t is the index of iteration and $\eta(t) = \frac{\eta(0)}{1+10^{-4} \times t}$ is a variable learning rate scheme with an initialization $\eta(0) = 0.0001$. We empirically set the maximum iteration number to 100,000, the hyper parameter λ to 3.

At the testing stage, let the probabilistic prediction given by both DCNN components be denoted by $\mathbf{P}^{(i)} = (p_1^{(i)}, p_2^{(i)}, \dots, p_K^{(i)})$, $i=1, 2$. The corresponding class label given by the SDL model is

$$\underset{j}{\operatorname{argmax}} \left\{ \sum_{i=1}^2 p_1^{(i)}, \dots, \sum_{i=1}^2 p_j^{(i)}, \dots, \sum_{i=1}^2 p_K^{(i)} \right\}\tag{6}$$

4 Results

Comparison to the Ensemble Learning: Figure 3 shows the receiver operating characteristic (ROC) curves and area under the ROC curve (AUC value) obtained by applying the ensemble of two ResNet-50 (ResNet-50²) and proposed SDL model without DA to the test set, respectively. It reveals that our SDL model (red curves) outperforms ResNet-50² (blue curves). More comprehensively, we give the average precision (AP), classification accuracy (Acc) and

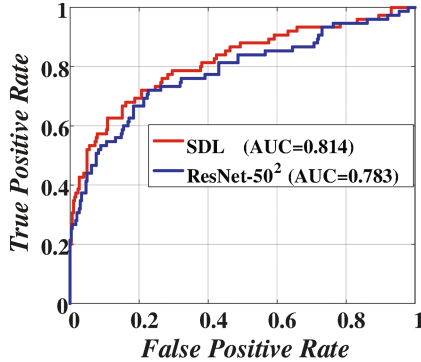


Fig. 3. ROC curves and AUC values of the proposed SDL model and ResNet-50².

Table 1. Performance of ResNet-50, ResNet-50² and SDL with or without DA.

Methods	DA	AP	Acc	AUC
ResNet-50	N	0.6102	0.8496	0.7742
	Y	0.6224	0.8522	0.7829
ResNet-50 ²	N	0.6115	0.8443	0.7826
	Y	0.6308	0.8549	0.7968
SDL	N	0.6536	0.8522	0.8139
	Y	0.6644	0.8575	0.8179

AUC value of ResNet-50, ResNet-50² and the proposed SDL model on the test set with or without DA in Table 1. It shows that SDL performs steadily better than ResNet-50 and ResNet-50² regarding three evaluation metrics no matter using or not using DA. It clearly demonstrates that the synergic learning strategy makes a big contribution to higher performance of the SDL model, compared with ResNet-50² without synergic learning.

Comparison to the State-of-the-Art Methods: Table 2 shows the performance of the proposed SDL model and the top five challenge records ¹, which were ranked based on AP , a more suitable evaluation metric for unbalanced binary classification [13]. Among these six solutions, our SDL model achieved the highest AP , highest Acc and second highest AUC . The 1st place method [4] leveraged a segmentation network to extract lesion objects based on the segmented results, for helping the classification network focus on more representative and specific regions. Without using segmentation, the SDL model still achieved a higher performance in skin lesion classification by using synergic learning strategy.

¹ <https://challenge.kitware.com/#phase/5667455bcad3a56fac786791>.

Table 2. Performance of the proposed SDL model and the top five challenge records in the leaderboard. Note that AP is the only evaluation metric, according to which all participants were ranked.

Methods	$AP(^*)$	Acc	AUC
SDL	0.664	0.858	0.818
CUMED [4]	0.637	0.855	0.804
GTDL	0.619	0.813	0.802
BF-TB	0.598	0.834	0.826
ThrunLab	0.563	0.786	0.796
Jordan Yap	0.559	0.844	0.775

5 Discussion

Stability Interval of Hyper Parameter λ : The hyper parameter λ is important in the propose SDL model. Figure 4 shows the variation of the AP of SDL over λ . It reveals that, as λ increases, the AP of SDL monotonically increases when λ is less than 3 and monotonically decreases otherwise in the validation set. Therefore, we suggest setting the value of λ to 3 for better performance.

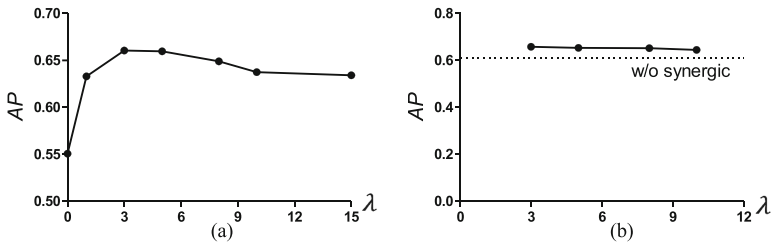


Fig. 4. Variation of the AP of our SDL model (without DA) on the validation set (a) and test set (b) over λ .

Synergic Learning and Ensemble Learning: The proposed SDL model can be easily extended to the SDL^n model, in which there are n DCNN components and C_n^2 synergic networks. Different from the ensemble learning, the synergic learning enables n DCNNs to mutually learn from each other. Hence, the SDL^n model benefits from not only the ensemble of multiple DCNNs, but also the synergic learning strategy. We plotted the AP and relative time-cost (TC) of the SDL^n model versus the number of DCNN components in Fig. 5. The relative TC is defined as the ratio between the training time of SDL^n and the training time of the single ResNet-50. It shows that, with the increase of DCNN components, the TC grows significantly and monotonically, whereas the improvement of AP

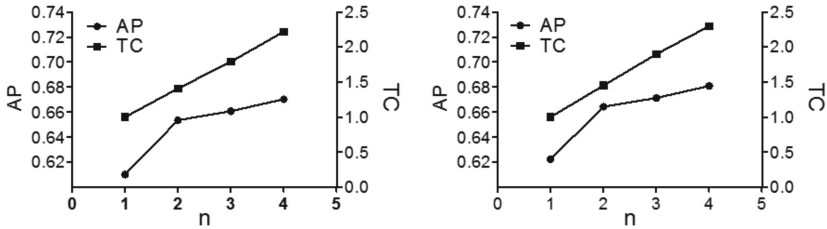


Fig. 5. Performance-time curve of the SDL^n model in the test set (left: without DA, right: with DA) when n changes from 1 to 4 ($n = 1$ represents single ResNet-50).

is first sharply and then becomes slowly when using more than two DCNNs. Therefore, taking the computational complexity into consideration, we suggest using the SDL^2 and SDL^3 models.

6 Conclusion

In this paper, we propose a synergic deep learning (SDL) model to address the challenge caused by the intra-class variation and inter-class similarity for skin lesion classification. The SDL model simultaneously uses dual DCNNs with a synergic network to enable dual DCNNs to mutually learn from each other. Our results on the ISIC 2016 Skin Lesion Classification dataset show that the proposed SDL model achieves the state-of-the-art performance in the skin lesion classification task.

Acknowledgments. This work was supported in part by the National Natural Science Foundation of China under Grants 61771397 and 61471297.

References

1. Siegel, R.L., Miller, K.D., Jemal, A.: Cancer statistics, 2016. *CA Cancer J. Clin.* **66**(1), 7–30 (2016)
2. Gutman, D., et al.: Skin Lesion Analysis toward Melanoma Detection: A Challenge at the International Symposium on Biomedical Imaging (ISBI) (2016). [arXiv:1605.01397](https://arxiv.org/abs/1605.01397)
3. Ge, Z., Demyanov, S., Chakravorty, R., Bowling, A., Garnavi, R.: Skin disease recognition using deep saliency features and multimodal learning of dermoscopy and clinical images. In: Descoteaux, M., Maier-Hein, L., Franz, A., Jannin, P., Collins, D.L., Duchesne, S. (eds.) *MICCAI 2017*. LNCS, vol. 10435, pp. 250–258. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-66179-7_29
4. Yu, L., Chen, H., Dou, Q., Qin, J., Heng, P.A.: Automated melanoma recognition in dermoscopy images via very deep residual networks. *IEEE Trans. Med. Imaging* **36**(4), 994–1004 (2017)
5. Esteva, A.: Dermatologist-level classification of skin cancer with deep neural networks. *Nat. Res.* **542**(7639), 115–118 (2017)

6. Weese, J., Lorenz, C.: Four challenges in medical image analysis from an industrial perspective. *Med. Image Anal.* **33**, 44–49 (2016)
7. Oquab, M., Bottou, L., Laptev, I., Sivic, J.: Learning and transferring mid-level image representations using convolutional neural networks. In: *CVPR*, pp. 1717–1724 (2014)
8. Zhou, Z., Shin, J., Zhang, L., Liang J.: Fine-tuning convolutional neural networks for biomedical image analysis: actively and incrementally. In: *CVPR*, pp. 4761–4772 (2017)
9. Xie, Y., Xia, Y., Zhang, J., Feng, D.D., Fulham, M., Cai, W.: Transferable multi-model ensemble for benign-malignant lung nodule classification on chest CT. In: Descoteaux, M., Maier-Hein, L., Franz, A., Jannin, P., Collins, D.L., Duchesne, S. (eds.) *MICCAI 2017*. LNCS, vol. 10435, pp. 656–664. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-66179-7_75
10. Song, Y., et al.: Large margin local estimate with applications to medical image classification. *IEEE Trans. Med. Imaging* **34**(6), 1362–1377 (2015)
11. Bengio, Y., Collobert, R., Weston, J.: Curriculum learning. In: *ICML*, pp. 41–48 (2016)
12. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: *CVPR*, pp. 770–778 (2016)
13. Yuan, Y., Su, W., Zhu, M.: Threshold-free measures for assessing the performance of medical screening tests. *Front. Public Health* **3**, 57 (2015)