ISIC2018 Challenge
Skin Lesion Segmentation, Attribute Detection and Classification via Deep Learning

Tian-Ruei Kuan, Liang-Yu Fan Chiang, Tzu-Shin Lin, and Shang-Hong Lai
Dept. of Computer Science, National Tsing Hua University, Taiwan

ABSTRACT

In this paper, we present the deep learning based methods to resolve the three tasks in ISIC 2018 challenge: *skin lesion analysis towards Melanoma diseases*. The proposed deep learning methods are based on training the global convolutional network models with the datasets provided by this challenge. We describe how the models are trained for these three tracks. The experimental validation of the trained models is shown with some discussions.

1. INTRODUCTION

Skin cancer has become a commonly encountered disease in recent years. To improve the skin cancer detection accuracy and enhance the associated disease treatment, International Skin Imaging Collaboration (ISIC) held this challenge [1] and provide a large and publicly available image dataset collection for skin cancer research.

This challenge includes the following three tasks:

Task 1 — lesion segmentation: to provide automated predictions of lesion segmentation boundaries from dermoscopic images.

Task 2 — lesion attribute detection: to automatically predict the locations of dermoscopic attributes from dermoscopic images. The classes include: pigment network, negative network, streaks, milia-like cysts, globules (including dots).

Task 3 — disease classification: to classify disease category from dermoscopic images. The classes include: Melanoma, Melanocytic nevus, Basal cell carcinoma, Actinic keratosis / Bowen's disease, Benign keratosis, Dermatofibroma, Vascular lesion.

We propose to train convolutional neural network (CNN) models for each of the three tasks with the training datasets provided by in the challenge. The details of the model training and model validation results are described in the subsequent sections.

2. PROPOSED METHOD

2.1 Image Pre-processing

The dermoscopic images usually contain hairs, which can be considered as noises that may degrade the model prediction accurate. Before the model training and, preprocessing the images to remove the hair regions may be helpful to the model prediction accuracy.

In this work, we propose to preprocess the dermoscopic images to remove hairs in the images. By using adaptive thresholding to obtain relatively darker and lighter areas in the input image, we can filter out hair and scars from other parts. Then we separate hairs from scar based on the different characteristics in the image appearance and replace hair regions with blurred images. Last, we apply the color constancy [2] for each image.

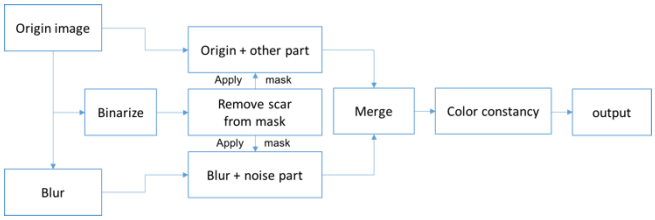

Fig 1. Flow chart of image pre-processing process

2.2 Task 1: Lesion Boundary Segmentation

The goal of task 1 is to predict the boundary of skin lesion from dermoscopic images. Each image contains only a single contiguous region. The goal metric is computed using a threshold Jaccard index. The final score is taking the mean Jaccard index of each image but images with scores less than 0.65 are count as 0.

Dataset

The ISIC 2018 Challenge provides 2594 dermoscopic images with various sizes for task1 and task2. We divide the images into 2494 training data and 100 validation data. The preprocessing described is first applied to get rid of the hair and other obstacles. The images are then resized to a fixed size of 256 × 256.

Model Architecture and Training

We use the global convolutional network (GCN) and the same pipeline described in [3]. GCN uses a combination of 1 × k + k × 1 and k × 1 + 1 × k convolution, which can have a better receptive filed without increasing the number of parameters dramatically. Since the goal of this task is to identify the boundary, larger receptive filed can give the

model more global information and thus performs better. The kernel size parameter k of the model is set to 7 and the dimension is reduced to 1 to match the number of classes of the dataset. Sigmoid nonlinearity is applied at the output. Since the challenge uses Jaccard index as the goal metric, we use Jaccard loss similar to [4] as our loss function.

$$\text{loss} = 1 - \frac{\sum y_{ij} t_{ij}}{\sum y_{ij} + \sum t_{ij} - \sum y_{ij} t_{ij}}$$

where $y_{ij} \in [0, 1]$ and $t_{ij} \in \{0, 1\}$ are the output and target for each pixel at position (i, j).

Data argumentation of random horizontal flip and vertical flip are used. The model is trained via Stochastic Gradient Descent (SGD) optimizer with the learning rate set to 0.001 and drop the learning rate by a factor of 0.1 once the loss reaches plateau twice. The batch size is set to 6 and the network is trained about 100 epochs.

2.3 Task 2: Lesion Attribute Detection

In task2, we need to identify 5 kinds of dermoscopic attributes: pigment network, negative network, streaks, milia-like cysts, and globules. We use one binary mask for each type. The goal metric is Jaccard index.

Dataset

We separate the 2594 images into training and validation data as described in task1. The images are resized to 512 × 512 without preprocessing. Other than the 5 given classes, we add a no-finding class. That is, when a pixel does not belong to any of the 5 classes, it belongs to the no-finding class as depicted in Figure 2. By adding the no-finding class, we can compute softmax directly from the output.

Fig 2. 5 attribute classes (left 1-5) and the no-finding class (right)

Model Architecture and Training

The same GCN [3] model is used with kernel size k = 7. The dimension is set to 6 to match the 5 classes and the no-fining class. Softmax is applied at the output.

The difficulty we encountered for this task is that the data is highly unbalanced. Most images have no positive instances for the attributes except the pigment network. As the result, we tried different loss functions. We found that a combination of weighted cross entropy loss and Jaccard loss works the best. The weighted cross entropy loss is given as follows:

$$loss_{WCE} = -\sum_{c} w_c t_c log\, y_c$$

where w_c , $t_c \in \{0, 1\}$, and $y_c \in [0, 1]$ are weight, target, and output for class c. The weight is 0.2 for the no-fining class and 1 for the other classes. Jaccard loss is calculated without the no-finding class. The final loss is then summed up with parameter α =0.8.

$$loss = \alpha loss_{Jaccard} + (1 - \alpha) loss_{WCE}$$

The model is trained via Stochastic Gradient Descent (SGD) optimizer with the learning rate of 0.0001. The batch size is 6 and the model is trained about 80 epochs.

2.4 Task 3: Lesion Diagnosis

The goal of Task 3 is to generate predictions from dermoscopic images for the following 7 disease categories: Melanoma (MEL), Melanocytic nevus (NV), Basal cell carcinoma (BCC), Actinic keratosis (AKIEC), Benign keratosis (BKL), Dermatofibroma (DF), Vascular lesion (VASC). The response data should include diagnosis confidence, which lies in a closed interval [0.0, 1.0], for each class. The evaluation is computed using a normalized multi-class accuracy metric.

Dataset

In Task 3: The official provides 10015 images [5] and 1 ground truth response CSV file. The numbers of images for each class are: 1113 for MEL, 6705 for NV, 514 for BCC, 327 for AKIEC, 1099 for BKL, 115 for DF, 142 for VASC . We divide this dataset into 8012 training samples and 2004 validation samples. The ratio for training data to validation data is approximately 4:1. The images are preprocessed to eliminate the hair and leave us more clear lesions. Then the images are resized to 224 × 224.

Table 1. Validation scores of different tasks

Class	MEL	NV	BCC	AKIEC	BKL	DF	VASC
#images	1113	6705	514	327	1099	115	142

Model Architecture and Training

We use ResNet-50 [6] as our training model. ResNet introduces the concept of Residual function to solve the problem that using deeper networks will degrade the performance of the model. It uses kernel size of 1 or 3 and down samples with CNN layers of stride 2. We add our fully-connected layer at the end of the network to classify 7 classes as needed. Using Softmax to calculate the probability for each class. Cross entropy is used as our loss function.

$$\text{loss}(x, class) = -\log\left(\frac{\exp(x[class])}{\sum_j \exp(x[j])}\right)$$

where *x* is the output of the model and *class* is the index of ground truth class.

We use SGD as our optimizer with a learning rate of 0.001 The batch size is set to 8 and trained about 300 epochs.

3. EXPERIMENTAL RESULTS

Some examples of the image preprocessing results are shown in Figure 3. We can see most hair regions are considerably removed or reduced in the images. In most cases, short hairs and scars are difficult to distinguish, since they both have small surface and low height/width ratio with dark color, thus making the classifier difficult to separate them.

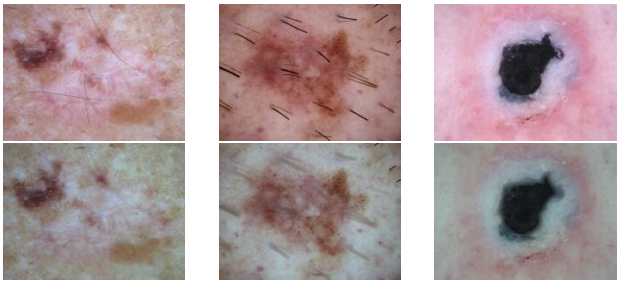

Fig 3. examples before/after pre-processing
(Top: input; Bottom: output)

The scores of all 3 tasks are shown in Table 1. For task 1 and task 2, the scores were calculated after applying a threshold of 0.5.

Table 2. Validation scores of different tasks
(Task1: threshold Jaccard index, Task2: Jaccard index, Task3: normalized multi-class accuracy)

Task Type	Task1	Task2	Task3
Score	0.727	0.379	0.699

The results of task 1 are shown in Figure 4. Most of the models can identify the boundary of the lesion. However, in some cases it may fail to identify the whole lesion and instead only recognize some part of it.

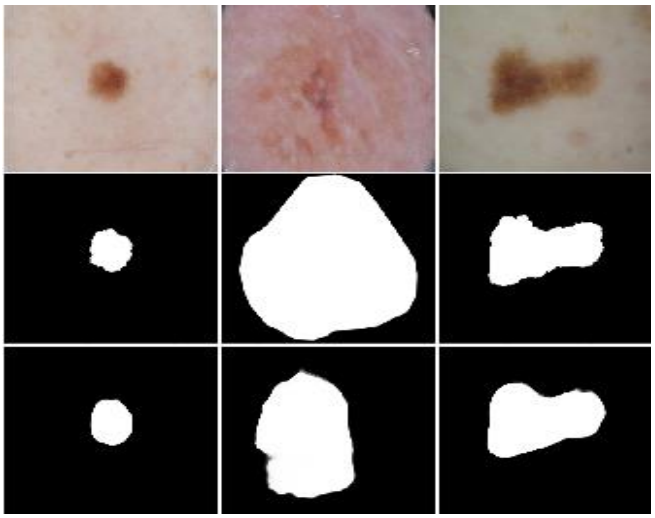

Fig 4. Sample result images for task1
(Top: input. Middle: ground truth. Bottom: output.)

For task 2, only the pigment network attribute may be poorly identified and other attributes mostly cannot have positive outputs. This is due to the fact that the data is highly imbalanced and we are not able to train the model reliably using the data given.

Fig 5. Sample result images for task2(From left to right: input, milia-like cysts ground truth, milia-like cysts output, pigment network ground truth, and pigment network output)

In task 3, the AUC of ROC of our best result is shown in table below.

Table 3. AUC result of each class in task 3

	MEL	NV	BCC	AKIEC	BKL	DF	VASC
AUC	0.700	0.795	0.930	0.936	0.778	0.949	0.917
Final AUC: 0.8579							

4. CONCLUSION

For image pre-processing, sometimes the hair regions cannot be removed very nicely. This may influence the accuracy of our CNN model. The current hair removal method used in this challenge is based on heuristics. In the future, developing a better hair removal method may improve the skin cancer detection and classification accuracy.

Data imbalance is a big problem that greatly influences the result in task 2. To solve this problem, applying suitable data augmentation technique could help overcome this problem and enhance the model accuracy.

5. REFERENCES

[1] Noel C. F. Codella, David Gutman, M. Emre Celebi, Brian Helba, Michael A. Marchetti, Stephen W. Dusza, Aadi Kalloo, Konstantinos Liopyris, Nabin Mishra, Harald Kittler, Allan Halpern: "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)", 2017; arXiv:1710.05006.

[2] Catarina Barata, M. Emre Celebi, Senior Member, IEEE, and Jorge S. Marques, Improving Dermoscopy Image Classification Using Color Constancy

[3] C. Peng, X. Zhang, G. Yu, G. Luo, and J. Sun. Large kernel matters–improve semantic segmentation by global convolutional network. *arXiv:1703.02719*, 2017.

[4] Yading Yuan. Automatic skin lesion segmentation with fully convolutional-deconvolutional networks. arXiv:1703.05165

[5] Tschandl P., Rosendahl C. & Kittler H. The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions. Sci. Data 5, 180161 doi.10.1038/sdata.2018.161 (2018)

[6] Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun. Deep Residual Learning for Image Recognition. *arXiv:1512.03385, 2015*