Skin Lesion Classification Using Deep Convolutional Neural Networks

Kaiqiang Ma[1] and Lingling Sun[2]

[1] Hangzhou Dianzi University, Hangzhou 310018, P. R. China
460996917@qq.com

Abstract. Abstract. Skin disease is one of the diseases which has the fastest morbidity in recent years, among them, malignant melanoma jeopardizes the most and its lethality is relatively high. It is very hard to diagnose, even the experienced dermatologist can make mistakes. Our team participated in the part 3 of ISIC, with the aim to increase the accuracy rate of identifying the skin disease with the network of CNN. The main problem we encountered is the imbalance of data set, the high resemblance among the different categories and some samples are hard to identify. We solve the problems of imbalance through down-sampling, use focal loss to learn the characteristics of samples and obtain the better between-class distance, with the combination of characteristic of different models to get a better result. Currently the result of single model in our divided validation set is accuracy:0.95，balanced_accuracy:0.90.

Keywords: under sample, focal loss, model feature ensemble.

1 Introduction

We participated in the third part of the ISIC competition. The defined task of the Part3 of the ISIC 2018 Skin Lesion Analysis Towards Melanoma. Detection. [3] was to classify the dermoscopic images into 7 classes: Nevus, Dermatofibroma, Melanoma, Pigmented Bowen's, Pigmented Benign Keratoses, Basal Cell Carcinoma, Vascular. The data set is provided by the ISIC official and contains a total of 10015 images and a CSV file of supplemental information including diagnosis confirm type. Our data was extracted from the "ISIC 2018: Skin Lesion Analysis Towards Melanoma Detection" grand challenge datasets [1][2]. The CSV file helps to better divide the data into train and validation set. Predicted responses are scored using a normalized multi-class accuracy metric (balanced across categories) which is the same as average recall.

In this report, we introduce our methods, and its results of different methods in our divided validation set and online validation set. But the online validation score results are not meant to be indicative of performance on the test dataset because of its small size.

2 Methods

2.1 Deep Residual Networks and Inception-v3

We both tried ResNet [4] and Inception-v3 [5]. In our experiments the balanced accuracy with Inception-v3 is higher than the balanced accuracy with ResNet. We used Kears and Pytorch library, the model of Pytorch achieved a higher score.

2.2 Under-sampling

Under-sampling [6, 7] is a technique that manipulate sample at the data level, in order to alter the training data distribution to have a good classifier. In deep learning, under-sampling is preferred to over-sampling, because over-sampling can easily introduce overfitting risks.

2.3 Focal loss

Focal loss [8] is submitted by Kaiming He's team. As their paper [8] demonstrated, the focal loss trains a highly accurate dense object detectors in the presence of vast numbers of easy background examples. Focal loss is an improvement of cross entropy, which adds a factor $\alpha_t(1-p_t)^\gamma$. They defined focal loss as:

$$FL(p_t) = -\alpha_t(1-p_t)^\gamma \log(p_t) \tag{1}$$

y specifies the ground-truth class, p is the model's estimated probability for the class, and p_t is defined as:

$$p_t = \begin{cases} p & if\ y = 1 \\ 1-p & otherwise \end{cases} \tag{2}$$

Focal loss can put more focus on misclassified examples, because of α_t , which is similar to cost-sensitive learning. It can also put more focus on hard examples, because of the p_t. We made a comparison of two classes, the model with focal loss performs slightly better, but it needs to adjust parameters carefully.

2.4 Multi-layers feature maps learning

Most classifiers use the feature map of the last convolutional layer, it performs not badly, but it ignores the feature maps of the previous layers, the output of different layers represent different levels of feature maps. In our paper we concatenate on different feature maps to produce an element vector, then put this vector into a classifier. With this method, we can get a more comprehensive feature map. The model architecture is demonstrated in Fig1

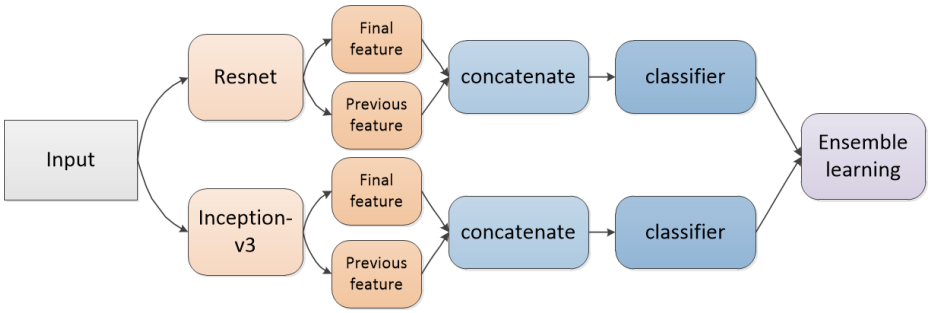

Fig. 1. Model architecture

3 Results

We show our evaluation results. Our divided validation set (995 samples) is composed of ten percentage of the training set. The online validation set was extracted from the "ISIC 2018: Skin Lesion Analysis Towards Melanoma Detection" grand challenge datasets [1][2].

In Table 1, we show the best score of single model in divided validation set. Because this validation is big enough.

Table 1. Recall of each class

method	MEL	NV	BCC	AKIEC	BKL	DF	VASC	MEAN
Inception-v3	0.829	0.978	0.941	0.75	0.990	0.909	0.923	0.902

In Table 2, we show the score of ensemble learning. Due to the small online validation set, there is a great difference between the results of different models.

Table 2. Accuracy

method	balanced accuracy
Ensemble learning	0.940

When we compared the results of models in different validation sets, we found that some models that performed well on the divided validation set did not necessarily perform well on online validation sets. One of the reasons may be that the validation set is too small.

4 Conclusions and Future Work

In this article, we described several methods and show the results of our models in two validation sets. And we adopt multi-feature maps learning, the approach allows to

combine feature maps of different layers. In our experiments, Inception-v3 works well, and under-sampling has its own advantages.

As described above, due to the small size of validation set, it can have a great impact on balanced accuracy, an official test set can help us modify our classification better.

References

1. 1. 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)
2. 2. 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.
3. ISIC 2018, https://challenge2018.isic-archive.com/, last accessed 2018/6/20.
4. K. He, X. Zhang, S. Ren: Deep residual learning for image recognition, CVPR 2016, 770-778. IEEE, Las Vegas (2016).
5. C. Szegedy, V.Vanhoucke, S. Ioffe: Rethinking the inception architecture for computer vision, CVPR 2016, 2818-2826. IEEE, Las Vegas (2016).
6. C. Drummond and R. C. Holte. C4.5: class imbalance, and cost sensitivity: why under-sampling beats over-sampling, ICMLW 2003, 11: 1-8. Washington DC (2003)
7. Z.-H. Zhou and X.-Y. Liu: Training cost-sensitive neural networks with methods addressing the class imbalance problem, TKDE 2006 , 18(1):63-77,2006. IEEE(2016).
8. T.-Y. Lin, P. Goyal, R. Girshick: Focal loss for dense object detection. arXiv:1708.02002, 2017.