Informatics in Medicine Unlocked 12 (2018) 67–74

Contents lists available at ScienceDirect

Informatics in Medicine Unlocked

journal homepage: www.elsevier.com/locate/imu

Multi-Channel-ResNet: An integration framework towards skin lesion analysis

Songtao Guo, Zhouwang Yang*

University of Science and Technology of China, China

ARTICLE INFO

Keywords:
Skin diseases
Medical image classification
Multi-Channel-ResNet
Model ensemble
Deep convolutional neural network

ABSTRACT

This paper proposes a feasible framework towards skin lesion analysis, named Multi-Channel-ResNet. The basic idea is to assemble multiple residual neural networks (ResNets), in which the training data has been pretreated with different methods. For different situations with practical applications, we put forward two training methods. Our method performs better than a single ResNet or a simple ensemble of ResNets. The validity of the framework is verified on two data sets: dermoscopic images and skin surface photos. For dermoscopic images, we completed the third part of a public competition, called "ISIC 2017: Skin Lesion Analysis Towards Melanoma Detection". The metric is the mean value of the area under the curve (AUC) for the melanoma and seborrheic keratosis classifications. Our framework achieves a result of 0.917 on the test set, which is 0.046 higher than a single ResNet. For skin surface photos, we collected images of four diseases, including eczema, heatrash, subitum, and varicella. The framework achieves 82.4% accuracy on the test set, which is 3% higher than the baseline. This implies that the proposed framework is applicable in practice and achieves excellent performance.

1. Introduction

Skin is the largest organ of the human body. It protects the body, helps perceive the outside world, and has many other important functions. Skin diseases are one of the most common forms of infections among people and cause great discomfort [1]. There are more than 2000 kinds of skin diseases, some of which are very similar. So, only experienced doctors can make accurate diagnoses in the early stages of diseases; however, the number of such experts is far from enough, and the cultivation of such experts demands extensive resources. Thus, it is difficult to diagnose a skin disease in a timely and convenient manner.

Developing a computer aided system increases the convenience of diagnosing certain diseases. As early as 1987, there was a study on the auxiliary diagnosis of skin cancer based on images [2]. Most of the traditional studies were about image processing, edge detection and image segmentation, and feature selection and extraction [3–5]. In recent years, with the rapid development of deep learning, great breakthroughs have been achieved in many fields, such as image classification [6,7] and speech recognition [8], that surpass the capabilities of human beings. At the same time, with the development and popularization of medical imaging technology, more and more image data can be used for clinical diagnosis and medical research. If deep learning technology is applied to medical imagery to assist doctors in disease diagnosis, doctors' efficiency and accuracy would improve.

There have been a number of explorations about deep learning in skin lesion analysis. Some of them directly apply the deep convolutional neural network (DCNN) to accomplish the classification task [9], some combine deep learning with sparse coding and support vector machines (SVM) [10], and some simply integrate the DCNNs (for example, by averaging the results of multiple models) [11]. One of the great advantages of deep learning, in contrast to traditional methods, is that it automatically learns high-level features in images. The learned features are usually better than the features selected by researchers in classification tasks. However, deep learning has its shortcomings. For example, further improvement of deep learning models is difficult because such a process requires big datasets. Unfortunately, the collection of data is not easy. One major reason is that the calibration of medical images is rather difficult and requires seasoned experts. Often these experts cannot even guarantee that they are completely correct, so validation from multiple experts is needed. And the designed diagnosis system usually has certain requirements for the specifications of the images.

In this paper, we propose a feasible framework based on deep learning for medical image analysis, named Multi-Channel-ResNet. It can be seen as an integration of the DCNNs that uses residual neural network (ResNet) [12]. Due to the robustness of DCNN, the required image specifications can be properly relaxed. After applying different pre-processing methods, images are put into different ResNets, respectively, and the feature layers are concatenated to get the final

* Corresponding author.
E-mail address: yangzw@ustc.edu.cn (Z. Yang).

https://doi.org/10.1016/j.imu.2018.06.006
Received 9 April 2018; Received in revised form 22 June 2018; Accepted 24 June 2018
Available online 27 June 2018
2352-9148/

prediction probability. There are two different training methods. The first method divides the Multi-Channel-ResNet into two parts for training purposes. In this case, we train the component ResNet models first, and then use an artificial neural network (ANN) to connect the concatenated layer with the output layer and train the ANN. The second method trains the Multi-Channel-ResNet as one model. In this case, we connect the concatenated layer and output layer directly. For different data sets, we adopt various pre-processing methods and flexibly apply different training methods. These varying pre-processing methods can be seen as a method of data augmentation to some extent. In our framework, we take advantage of ResNet's excellent learning ability. Conventional data augmentation methods are also applied, such as rotation, shift, flip, etc. Along with transfer learning [13], they act as methods to confront the lack of data.

The framework is applied to two kinds of medical images in this paper: dermoscopic images and skin surface photos. The data are all reliable, but insufficient. For the dermoscopic images, we classify them as melanoma, nevus, or seborrheic keratosis. Skin cancer is a major public health problem, and melanoma is the deadliest form of skin cancer [14]. Thus, early detection of melanoma is of great significance. For the skin surface photos, we categorize them as eczema, heatrash, subitum, or varicella. The rash images depict four skin diseases that are common amongst infants and young children. Although these diseases occur frequently, they are difficult to distinguish due to their similar appearance. Thus, providing early diagnoses is clearly in the interest of both children and parents. It is therefore evident that the two applications explored in this paper have strong practical applications.

1.1. Our contributions can be summarized as follows

- 1. We propose a feasible framework for skin disease diagnosis, which has been validated on two kinds of data, including dermoscopic images and skin surface photos. Similar frameworks can be extended to other diseases.
- 2. We provide an idea for DCNN integration and data augmentation, which highly improves the performance of DCNN. Also, the two proposed training methods adapt to different situations.

The remainder of the paper is structured as follows. Section 2 presents the details of our proposed framework, Section 3 includes experiments on dermoscopic images and skin surface photos, Section 4 presents a discussion, and finally, Section 5 draws the conclusions of this work.

2. Proposed framework

2.1. Multi-Channel-ResNet

Fig. 1 shows the structure of Multi-Channel-ResNet. As we can see, the input original image passes through several channels, each of which corresponds to a pre-processing method and a ResNet, hence the name, Multi-Channel-ResNet. Then, the fully connected layers, also known as feature layers, will be concatenated to obtain the predicted probability and the result. Note that there is an "Optional ANN" between "Concatenated Fully Connected Layer" and "Predicted Probability", which means that we can connect the two layers directly or through several layers of artificial neural network (ANN). This is generally based on training methods and will be explained in the next subsection.

For one input image, after each channel's fifty-layer ResNet (ResNet50), a 2048 dimensional feature vector is obtained, written as x_i, where i represents the i-th channel. If we assume that we are talking about a Three-Channel-ResNet, then the "Concatenated Fully Connected Layer" can be written as $x = (x_1, x_2, x_3)$, which has 6144 dimensions. If we assume that we have added a two-layer ANN to get the "Predicted Probability", then the hidden layer vector $h^{(1)}$ can be given by:

$$\boldsymbol{h}^{(1)} = \boldsymbol{g}^{(1)}(\boldsymbol{W}^{(1)}\boldsymbol{x} + \boldsymbol{b}^{(1)}) \tag{1}$$

where $\boldsymbol{W}$ represents the weights from the previous layer in the current hidden layer, $\boldsymbol{b}$ represents the bias, g is generally a nonlinear activation function, and all of the subscripts (1) represent the first hidden layer. There are a variety of options for g, such as Relu [15], Maxout [16], and so on. As for multiple classification problems, the activation function for the output layer is usually Softmax [17]:

$$y_i = softmax(a_i) = \frac{exp(a_i)}{\sum_{k=1}^{K} exp(a_k)} \tag{2}$$

where y_i represents the value of the output layer's i-th node, $\boldsymbol{a}$ represents the value of the output layer before activation (for the two-layer ANN, $\boldsymbol{a} = \boldsymbol{W}^{(2)}\boldsymbol{h}^{(1)} + \boldsymbol{b}^{(2)}$), and K is the total number of categories. Obviously, $\sum_{i=1}^{K} y_i = 1$, so every output y_i is regarded as the predicted probability of the corresponding category, and then, the one with the largest probability is selected as the predicted category.

We can adjust the number of channels according to the specific circumstances; by taking efficiency and effectiveness into account, generally two to four channels are suitable. For different types of data, we should apply different pre-processing methods based on the understanding of the images. The image pre-processing part can be regarded as a kind of augmentation of data, and the entire structure is actually an integration of ResNet. In actual use, the number of medical images is generally inadequate, so data augmentation supplements this insufficiency. Since ResNet performs well by itself already, the integration clearly achieves an even better performance.

2.2. Training methods

For different situations, we have designed two different training methods: Train-A and Train-B.

Train-A. Assume that we want to train a n-Channel-ResNet. The steps of Train-A are as follows:

- *Step*1: Train n ResNet models separately. The training data for each model are pretreated with different methods, and the labels are the same as the original data.
- *Step*2: To train an ANN, concatenate the fully connected layers of the n trained models as the input and use the labels of the corresponding original data as the output.
- *Step*3: Combine the previous two parts together and test the whole model. Return to Step1 if the results are not satisfying.

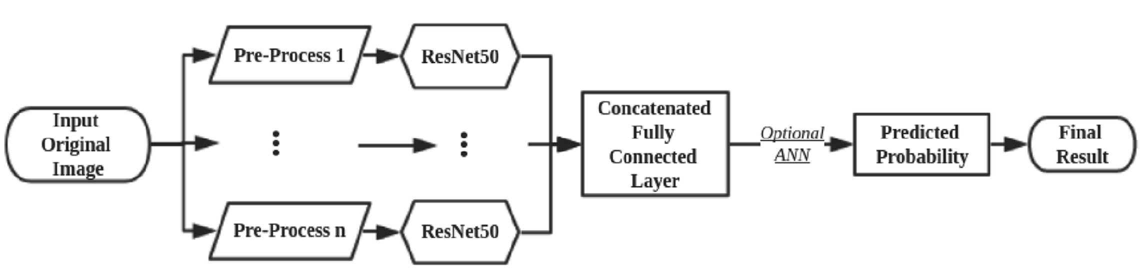

Fig. 1. Structure of multi-channel-ResNet.