Fusion of EfficientNet B0 and DenseNet121 *

Chuantao Xie

Shanghai University, Shanghai, 200444, China

1 Datasets

1.1 Data division

We divide the officially published training dataset into 9:1. Specifically, for each class, we extract 10% of the dataset as the verification set, and we treat the rest of the dataset set as the training set.

1.2 Data augmentation

We use the combined data augmentation method, which mainly includes four basic data augmentation methods: random rotation - 90 degrees to 90 degrees, horizontal flip, vertical flip, random clip and then resize. The combined data augmentation is carried out according to the above order. For each data augmentation method, the probability the mothed is selected is 0.5 . In general, there are 15 data augmentation methods.

2 Model

Two basic networks, eficientnet B0 and densenet121, are used for classified. As two lightweight networks with high classification recognition rate, eficientnet B0 and densenet121 are widely used in classification. The main reason for choosing these two networks is that when the two networks connect features, efficientnet B0 uses additive method for feature fusion, while densenet121 uses concat method for feature fusion. There are differences and complementarities between them.

3 Training

We train on a single GPU (NVIDIA Tesla P100) , with a batch size of 8 and Adam optimizera with a initial learning rate of 0.001. We use half of the sum of cross entropy of efficientnet B0 and densenet121 as the objective function of our optimization.If the accuracy of the model in the validation set does not increase after more than 30 cycles, the training will be stopped.During training, we use EfficientNet B0 pre-trained model and DenseNet121 pre-trained model, which

* Supported by organization x.

has been trained on the ImageNet dataset, to accelerate model convergence At the same time, for faster convergence, before training, we normalize the pixel values of each channel of the image, then the mean value of each channel is 0.5, and the standard deviation of each channel is 0.5.We can adopt the proposed method by replacing the layers after the main structure of the CNN, leaving the rest unchanged.The network structure before the global average pooling is retained, which a parallel network structure is connected with.One of the branches of the parallel network structure includes convolution, batch normalization, activation function, global average pooling and full connection, then the prediction of classes is carried out. Finally, the predicted results of the parallel network structure are concatenated, and then the full connection prediction and classification are carried out.The overall architecture of this network is shown in Fig.1.

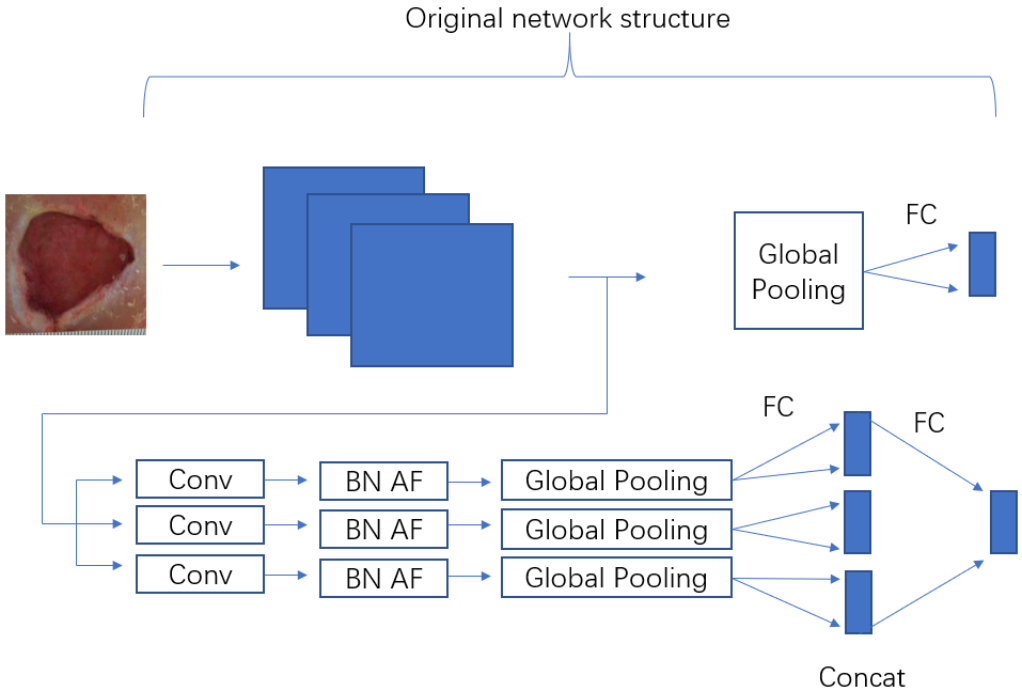

Fig. 1. Overall architecture of the proposed network

4 Test

For test dataset, we normalize the pixel values of each channel of the image, then the mean value of each channel is 0.5, and the standard deviation of each

channel is 0.5.The test images are input into two networks respectively, and the probability of each classification predicted by the two networks is averaged.

5 Other

email:ctxie@shu.edu.cn
link to github:https://github.com/ctxie/xie3.git