Hair Recognition Based on Multi-task Convolution Network

Xiaochao Guo, Liying Sui, Lu Shang, Dexue Zhang*, Tiantian Jia
College of Electronic Information Engineering, Shandong University of Science and Technology, Qingdao, 266590, China
gxc1228@yeah.net

Abstract. Aiming at the problem that the prior art is difficult to utilize the mutual relationship between labels to achieve efficient classification of hair style multi-labels, a hair styling classification method based on multi-task convolutional neural network is proposed. A multi-task joint learning model was constructed to try to realize the simultaneous recognition of hair shape and color. First, all the labels are jointly learned through the shared network layer, and the correlation between labels is automatically mined and utilized from the perspective of feature extraction. Then complete specific categories of learning tasks at different sub-network layers, thereby eliminating ambiguity in multi-label classifications. Finally, multiple classifiers are trained to achieve parallel prediction of all labels. The research shows that the proposed method can extract multiple features of the hair at the same time and directly classify and identify it. Compared with the single task network, the precision has obvious advantages.

1. Introduction
Hair styles belong to multi-label image data, labels can be divided into two categories: shape and color. There is some correlation between the two labels, for example, short hair is usually black. Therefore, in multi-label classification, the correlation between tags can provide more effective information for multi-label learning. At present, multi-label algorithms mainly try to use the correlation between tags as a prior knowledge to improve their classification performance.

Currently, most convolutional neural network structures only support single-label data. The most straightforward way to implement multi-label classification of hair styles using existing single-label networks is to train multiple models to convert multiple classifications into two classifications. However, if multiple labels are completely separated and feature extraction and classification are performed separately, the effective information provided by the associations between them may be ignored. Therefore, it is a key problem to realize hairstyle multi-label classification to automatically mine the correlation between tags and effectively apply it to the classification model.

In order to explore the correlation between multi-labels of hairstyles and achieve efficient and accurate classification of hairstyles, this paper uses deep convolution neural network to construct a multi-task joint learning model based on feature dependency, from the perspective of feature extraction and relevance learning, in order to achieve multi-label classification of hairstyles. In this paper, the two main attributes of hairstyle are the color and the shape, color are classified into black, golden, brown and grey. Choose four common attributes of straight hair, wavy hair, bald head and wearing a hat to judge the shape. The main attributes of hairstyle are identified and analyzed at the same time, this classification method also conforms to the judgment method of hairstyle in people's daily life.
2. Design of Multitask Network Structure
This section begins with the problem of multi-label classification of hairstyles, and elaborates the ideas and methods of constructing multi-tasking networks, there is a correlation between the recognition tasks of hair shape and color. This correlation means that two tasks can share some common features extracted from the convolutional layer. In fact, the article cited[1] showed that the risk of overfitting the shared parameters is smaller than overfitting the task-specific parameters. Intuitively speaking: the more tasks you learn at the same time, the more the model needs to find a feature that can capture all the tasks, the less likely it is to overfitting. Therefore, using a multi-tasking network as the structure of the model, by sharing representations between related tasks, the model can be better accomplished with the original task and avoid over-fitting.

For the specific tasks of the two classifications, the classification of colors only needs to judge the pixel values of large patches that satisfy a certain distribution law. For the classification of the shape of hair styles, especially for the texture of straight hair and wavy hair, the classifier not only considers the color, but also requires the classifier to summarize the local and large-range arrangement of the pixels. Therefore, classifying the shape of the hairstyle requires a deeper feature extraction. Since the five convolutional layers are large enough for hair color classification, an additional convolutional layer is added to the hair-shaped sub-network to extract the shape features to obtain deeper shape features.

Based on the above considerations, the network structure of the design is shown in Fig.1.

Fig.1 Structure of Multitask Network

In Fig.1, CL (1-5) is composed of batch standardization layer, convolution layer and pooling layer; CL6 is composed of batch standardization layer and convolution layer; FC (1-3, 5-7) is composed of batch standardization layer and full connection layer; FC (4, 8) has only one full connection layer.

The multi-task learning network proposed in this paper mainly consists of two parts. The first part is the shared layer of the dotted line. Through the hard parameter sharing, all tasks are learned at the same time, which facilitates the mining of deeper features. The second part is organized in the form of network branch. As shown in Fig.1, there are two network branches, which are used for color recognition and shape recognition training. Finally, all prediction results are output in parallel to realize simultaneous identification.

3. Multitasking Network Training
This section will elaborate on the training strategy, parameter setting and training process of the network in detail.

3.1. Parameter settings of Shared layer
The dotted line frame in Fig.1 is the structure of the shared layer. The parameter settings of this structure are shown in Table.1:

|      | Number of filter | Filter Size/stride | All the sublayer | Initialization of parameters |
|------|------------------|--------------------|------------------|-----------------------------|
| CL1  | 6                | 3×3/1              |                  | B+R+M                       | Xavier                      |
| CL2  | 12               | 3×3/1              |                  | B+R+M                       | Xavier                      |
| CL3  | 24               | 3×3/1              |                  | B+R+M                       | Xavier                      |
B represents the BN layer, R represents the activation function ReLU, and M represents the Maxpooling layer.

3.2. Parameter Settings of Two Subnetwork Structure

Two substructures in a multitasking network are shown in Fig.2:

![Fig.2](image_url)  
(a). Substructure of hair color classification; (b). Substructure of hair shape classification

The parameter settings for these two substructures are shown in Table.2:

| Sublayer | hidden units each layer | FilterSize/stride | All the Sublayers | Initialization of parameters |
|----------|-------------------------|-------------------|-------------------|-----------------------------|
| CL6      | 100                     | 2×2/1             | B+R               | Xavier                      |
| FC1/5    | 1000/1000               |                   | B+R               | Xavier                      |
| FC2/6    | 900/1000                |                   | B+R               | Xavier                      |
| FC3/7    | 600/1000                |                   | B+R               | Xavier                      |
| FC4/8    | 4/4                     |                   | R                 | Xavier                      |

3.3. Hairstyle Classification Training Process

The training process is mainly divided into forward propagation process and backward propagation process, forward propagation is the process of processing pictures and getting classification results in multi-task network, backward propagation is the process of adjusting parameters in multi-task network.

In the forward propagation process, the input original image passes through the convolution layer and the pooling layer, and several feature subgraphs are obtained and transformed into one-dimensional feature vectors. The feature vectors are input into the full connection layer, and further feature extraction and feature compression are carried out. Finally, the classification results are output, detailed formulas will not be repeated.

In the process of backward propagation, the output value of forward propagation is compared with the actual value, and a multitask loss function is established. The parameters of each layer are adjusted by random gradient descent method to make the output value closer to the actual value. The multitask loss function is shown in equation (1):

$$L = \omega_c \sum L_c + \omega_s \sum L_s$$  \hspace{1cm} (1)

$L_c$ is the color class loss function:

$$L_c(P_i, \{P_i^j\}) = -\log[P_i^iP_{\bar{i}} + (1-P_i^i)(1-P_{\bar{i}})]$$  \hspace{1cm} (2)

$L_s$ is the hair shape loss function:

$$L_s(P_j, \{P_j^s\}) = -\log[P_j^sP_{\bar{j}} + (1-P_j^s)(1-P_{\bar{j}})]$$  \hspace{1cm} (3)

In equation (1), $\omega_c$ and $\omega_s$ are weights that are set to eliminate the bias of the model, making the model more balanced. Add weights before the loss function of each task. When the accuracy of the two
classifications is very different, you can adjust the gradient of network update by adjusting the size of \( \omega_c \) and \( \omega_s \), slow down the fast convergence model, and accelerate the slow convergence model. Among them, \( P_{ci} \) and \( P_{sj} \) belong to the probability that the hair color and hair shape predicted by the model belong to the i-th and j-th categories respectively. \( P^* \) represents the actual probability. And only 1 and 0, \( P_{ci}^* = 0 \) means that the hair color does not belong to the first class, and \( P_{ci}^* = 1 \) means that it belongs to the i-th class.

The error between the predicted value and the given real value is calculated by loss function, and the back propagation algorithm is used to transmit the error layer by layer. The parameters of each layer are adjusted and updated by random gradient descent method, so that the predicted value of the network is closer to the real value, that is, the output of the last two full-connected layers is closer to the category and location information in the given label value.

4. Experiment Setup

4.1. Experimental data

In this paper, 11034 images that meet the requirements of this paper are selected from the CelebA data set, and 11034 additional images are generated by the image intensifier, totaling 22068 sample images. All the sample data are generated into training set, verification set and test set in a ratio of 6:2:2. The number of each label is shown in Table.3:

| Label          | Number | Label     | Number |
|----------------|--------|-----------|--------|
| Black hair     | 6722   | Bald      | 1103   |
| Blond hair     | 5342   | Straight  | 7988   |
| Brown hair     | 716    | Wavy hair | 12238  |
| Gray hair      | 998    | Wearing hat | 739   |

4.2. Lab environment

| Name           | parameter                        |
|----------------|----------------------------------|
| Computer system| Windows 10 Professional 64-bit   |
| Frame          | Tensorflow                       |
| Languages      | python                           |
| CPU            | Core i7-8700                     |
| GPU            | Nvidia GeForce GTX1060           |
| RAM            | 16GB                             |

5. Experimental Results and Analysis

To demonstrate the superiority of multitasking networks, a comparative experiment was designed: two single-task models were trained to identify hair color and hair shape. The two single task models are based on the multitasking network model in Fig.1, removing the unrelated subnetworks, and the parameter settings are the same as the multitasking network. The classification results are recorded and compared with the classification results of the multitasking network model designed in this paper to verify the classification performance of the multitasking network model.

The two single-task network models are respectively trained on the training set. The batch size of the input data during training is 30, the training set loss is output once every 50 batches, and the verification set loss is output once every 50 batches. Set epoch to 30, epoch is the number of traversal of the entire training set; after each epoch, a specific amount of data is randomly extracted from the test set to obtain and record the color/shape classification accuracy of the current network model pair. At the same time, the early stop algorithm is used to prevent over-fitting. The classification results of the two single-task network models are shown in Fig.3.
It can be observed that the single classification convolution network has a final classification accuracy of 93% for color, and the final classification accuracy for shape is 85%. There is a gap in accuracy and accuracy is not high. At the same time, the final difference between the training set and the verification set of the two models is large, and it can be judged that the model has over-fitting.

For the multitasking network model, the batch size of the input data during training is 30, the training set loss is output once every 50 batches, and the verification set loss is output once every 50 batches. Set epoch to 30, epoch is the number of traversal of the entire training set; multi-task network model classification results are shown in Fig.4:
As shown in Fig.4 (b), the classification of hair color and shape is 95% and 93%, respectively. Compared with the single task network, the accuracy is improved obviously, and the loss difference between the training set and the verification set is not obvious, which indicates that there is no serious over-fitting phenomenon. The experimental results show that the method of shared network can make effective use of label correlation and improve the performance of hairstyle classification, and only one model needs to be trained to recognize multiple labels.

Next, we use the visual analysis of feature extraction ability to evaluate the performance of multi-task network model. Here we choose saliency map method for feature analysis. Here are some saliency maps generated by our model.

As can be seen from the Fig.5, the high gradient pixels in the saliency map are very consistent with those in the original hair region, which further proves the ability of multi-task model to extract features accurately. Two very interesting phenomena are mentioned here:

1. As shown in Fig.5, there is a bald man in the image. The activation points of the corresponding salient map are significantly less than those of other people. Therefore, it is speculated that the classification mechanism of the model for baldness is to judge whether the person in the photo is bald or not, rather than to extract some features of baldness.

2. In saliency maps, activation points appear not only near the hair, but also in most areas of the whole face. It is speculated that the reason lies in the relationship between facial features and hairstyle. For example, if the color of the person in the picture is white, the likelihood that the hair color is golden will be significantly higher than that of black hair.

In conclusion, the performance of multi-task neural network is better than that of single-task neural network in hair color and hairstyle classification. It is proved that multi-task neural network may find...
some more essential and in-depth features in feature extraction for the classification of related tasks, which makes the classification effect of each task better.

6. Conclusion

This paper takes the correlation between multi-tasks as the starting point and constructs a multi-task joint learning model by using deep convolutional neural networks. At the same time, the automatic extraction of multiple labels of the hairstyle is realized, the dependence between the labels is enhanced, the correlation between the data is used to improve the classification performance, the end-to-end multi-label classification is realized, and the classification efficiency is improved. Moreover, the network structure proposed in this paper is simple, easy to expand and adjust, and can be used for classification tasks of more tags.

However, there are some shortcomings in this paper. Firstly, the data set we used for training is not created specifically for hairstyle classification, so there are some shortcomings in using it, such as: the total number of pictures is different; the number of pictures with different labels is too different. These data problems will affect the ultimate performance of the model, and although data enhancers are effective in mining deeper features, they have no effect on learning other new features that may be very useful for the model.

Secondly, due to the limited hardware performance, no more hyper-parameters are attempted to achieve better model performance.

Hair style is one of the important features of character recognition. The next work is to further improve the multi-task network model, and further study various types of hair style to get more types of hair style.

Acknowledgments

Special thanks to the teacher for their careful teaching and the help of classmates.

References

[1] Baxter, J. (1997). A Bayesian/information theoretic model of learning to learn via multiple task sampling. Machine Learning, 28,7–39.
[2] Lee J,Kim H,Kimn,R,et al . An approach for multi-label classification by directed acyclic graph with label correlation maximization[J].Information Sciences,2016,351(C):101-114.
[3] Wu J,Ye C,Sheng V S,et al. Active learning with label correlation exploration for multi-label image classification[J].Iet Computer Vision,2017,11(7):577-584.
[4] Li S L,Liu R,Liu H. Multi-Label Learning for Improved RBF Neural Networks[J].Computer Science,2015,42(4):316-320.
[5] Xu X D,Yao M H,Liu H W, et al. Pre-processing Method of Multi-label Classification Based on Knn[J]. Computer Science,2015.42(5).106-108.131.
[6] Zhao Limin Non-destructive testing of wall climbing robot body structure design and research [D].China university of Petroleum,2011.
[7] Yang Z H,Zhang D P, LI N M,Kernel False-Colour Transformation and Line Extraction for Fissured Tongue Image[J].Journal of Computer-Aided Design and Computer Graphics,2010,22(5):771-776.