Facial Age Estimation Method Based on Fusion Classification and Regression Model

Fengbing JIANG[1],Yu ZHANG[2],GuoLiang YANG[2]

[1]Science and Technology College of Gannan Normal University, GanZhou, JiangXi, China 341000
[2] School of Electrical Engineering and Automation, JiangXi University of Science and Technology, GanZhou, JiangXi, China 341000

Abstract. Due to the large individual differences in the facial features of each person and the fact that the age has a certain time sequence, the age estimation based on face images faces certain difficulties. This article proposes a method based on fusion classification and regression model: A classification model and a regression model are added to the convolutional neural network to train the network under the premise of sharing convolutional layer parameters. The classification of the age of the label is used to code the age distribution, and the age is regressed using the Euclidean distance. The final predicted value of the model is the average of the two. Experiments show that the effect of fusion classification and regression model is better than that of a single model, which improves the accuracy of age estimation.

1 Introduction

Age estimation based on face images has wide application prospects in pedestrian detection, face recognition, smart advertising, and security monitoring. Kwon and Lobo first studied face age estimation but could only approximate the age range. Geng et al. proposed a method of automatically estimating age based on the features included in the face image, which has a strong generalization ability for the absence of some age groups in the dataset. Lanitis et al. proposed a statistical model based on the appearance of human faces. The main task was to design a series of classifiers, transform the facial images into feature vectors, and then correspond the feature vectors to face age. Guo et al. extracted facial age features using different orientations and scales of Gabor filters based on the Bio-inspired Feature (BIF). Krizhevsky used an in-depth CNN structure in 2012 to obtain the best classification effect of ImageNet's large-scale visual recognition challenge competition at that time. The convolutional neural network structure has great potential under large-scale datasets[3]. By now, the application of convolutional neural networks has brought the accuracy of age estimation to a new height. Rothe et al. [5] adopted a migration learning and pre-training method to obtain an initial model by pre-training the IMDB-WIKI data set in the VGG-16 network.， which not only achieved an accuracy of 96.6% in the case of age segmentation and you can directly estimate the exact value of age. In order to solve the problem of convolutional neural network demands many datas, Hu et al [6] proposed a weak label method for age estimation, and achieved good experimental results. However, the number of images of the same individuals in the dataset used by Hu is relatively small and the age span is large, cannot intercommunicate, existence some limitations.

This paper presents a new method based on fusion classification and regression model. The experimental comparison results show that this method possesses some superiority in terms of age estimation accuracy.

2 Single label and label distribution

2.1 Single label

Assuming that the training sample set is $X = \{x_1, x_2, \cdots, x_n\}$， the label corresponding to each training sample is $Y = \{y_1, y_2, \cdots, y_n\}$. Divide the age label of face image training set $D = \{(x_1, y_1), (x_2, y_2), \cdots, (x_n, y_n),\}$ into a parts. Each label vector is represented as $y_m = [y_{m1}, y_{m2}, \cdots, y_{ma}]$. That is, each face image corresponds to a label vector. In the single label encoding method, the label position corresponding to the real age is recorded as 1 and the rest is 0.

2.2 Label Distribution

The encoding of the labels distribution means that the age labels of a face image are expressed using multiple ages, and the weight of each age-appropriate label is assigned. These weights are assigned using a probability distribution in accordance with a discrete Gaussian distribution. Using the Gaussian distribution to encode the age labels can be understood as the highest probability at the center age and the highest level of age description. As the distance from the center age

increasing, the probability decreases gradually, and the level of the description of the center age by the neighboring age label follows.

3 Motivation and Model Design

3.1 Ideas of the integration of classification and regression model

Multi-task learning is mainly to use the same training set to train certain tasks with certain correlations at the same time to achieve parallel processing of multiple tasks. For the problem of age estimation in this paper, the method of using the classification and regression fusion model is inspired by multi-task learning. However, not all problems can use the multi-task learning method. The premise of multi-task learning is multiple tasks must be relevant because unrelated tasks cause the network failure of converging during training. Age estimation is essentially a task in which the expected age value can be infinitely approximated to the actual age value, and the use of classification and regression methods can be regarded as tasks using different methods under the same goal. Age classification and age regression are highly relevant tasks, which meets the basic premise of multitask learning.

3.2 Age estimation of models incorporating classification and regression

The CNN model used in this paper proposes improvements based on the GoogleNet model. Compared with GoogLeNet and other popular CNN models, there are two main advantages: 1. A module named Inception is proposed. This module uses a smaller-scale convolution kernel instead of a larger-scale convolution kernel. That makes the number of this network parameters are reduced to a large extent; 2. Two Auxiliary Softmax layers are added for forward propagation to solve the problem of gradient disappearing during training.

This article uses a method of classification and regression models, adding a regression model to the network model. In the classification model, the age-label encoding method of label distribution is used, the classification layer uses the Softmax classifier, and the K-L divergence is used as the loss function. The regression model uses the Euclidean distance as the loss function for age regression. And the fusion classification and regression model is shown in Figure 1.

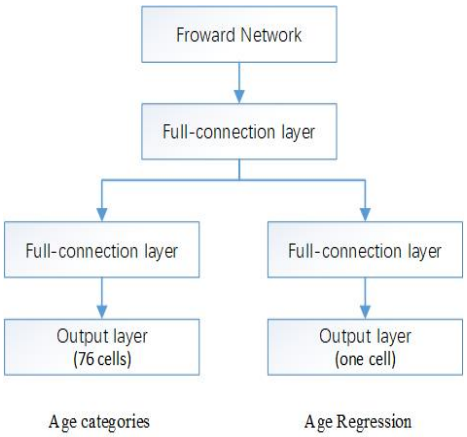

Figure 1: Classification and regression model fusion diagram

GoogLeNet used in this paper selected the Inception_v2 module [7] with higher precision and fewer parameters, and we proposed two improvements based on the original GoogLeNet: First, the original two auxiliary loss layers are deleted. Because the GoogLeNet network has a large number of parameters, the gradient disappears easily when training the network. The shallow parameters of the network are difficult to train, but we add BN to the network. The Batch Normalization layer can effectively solve the problem of the disappearance of gradients and reduce the risk of the network falling into overfitting; Second, add a full connection layer behind the last pooling layer of the network. After the full connection layer, add the fusion classification and regression model used in this paper, and do not use the Dropout operation commonly used in the full connection layer. Because this article needs to be pre-trained and fine-tuned before training the network, the full-connected layer often plays the role of "classifier" in the network, so the full-connected layer is indispensable. Figure 2 is the overall structure diagram of the CNN model used in this paper.

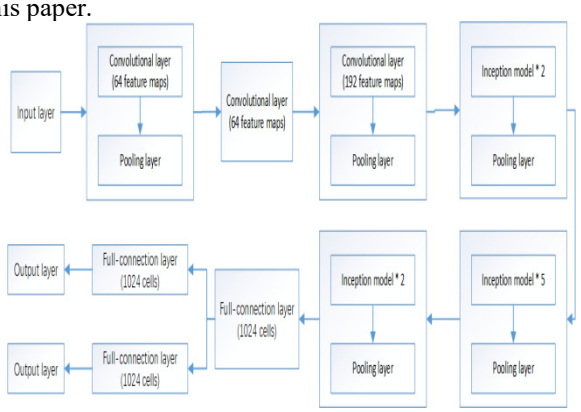

Figure 2: The overall structure of the CNN model

4 Experimental Details and Results Analysis

In order to verify the effectiveness of the fusion classification and regression model, the experimental results are evaluated on the basis of the fusion dataset which contains IMDB-WIKI dataset and the fusion dataset of MORPH and FG-NET. At the same time, the novel method was compared with representative face age

estimation algorithms in recent years in the MORPH dataset.

4.1 Experimental details

There are three parts in the training of convolution neural network. we adopted training strategy of from coarse to fine to reduce risk of over fitting. The IMDB-WIKI dataset what was filmed in non restrictive scenes, take MTCNN[4] to detect face and clip image. Figure 3 is a flow chart of the training network

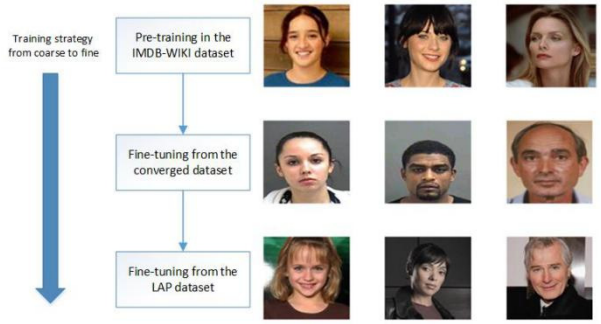

Figure 3. A flow chart of the training network.

We use Caffe framework to build CNN, that the initial learning rate is set at 0.01;Kinetic energy coefficient is 0.9; And the weight attenuation coefficient is 0.0005. We adopt various training strategy in some datasets. There are more details as follows:

(1)We pre-train the network to estimate age by sift 100 thousand face images from IMDB-WIKI dataset. There are 80 thousand images to pre-train and 20 thousand images to test, age range from 0 to 75. An age as a single label, adding a output layer of 76 dimensions at the network ending, adopting Softmax classifier and using cross-entropy as loss function, $\text{batch\_size} = 50$, iterating 200 thousand times, finally, an initial was obtained when the training was complete.

(2)Before using MORPH, FG-NET fusion and amplified dataset, the final output layer of the initial model is removed to replace it with the fusion model presented in this paper; then, use the updated dataset to fine-tuned the initial model, training sets having 50000 images and test sets having 13800 images, $\text{batch\_size}_{\text{training}} = 32$, $\text{batch\_size}_{\text{test}} = 50$, iterating 100 thousand times.

4.2 Experiment analysis

(1) Updated dataset for FG-NET and MORPH fusion

We take 5 methods what contain single label classification, label distribution, regression, single label + regression and label distribution + regression to verify the accuracy of the age estimates with the updated dataset. CS crave and MAE were used to make scientific evaluation.

Table 1. MAE value comparison table

Method	MAE
Single Label	3.49
Label Distribution	3.17
Regression	3.09
Single label + Regression	3.25
Label distribution + Regression	2.96

There are some information be shown in table 1 and figure 4:

(a)In single classification model, the accuracy rate of label distribution is better than single label.

(b)After merging regression model, the accuracy rate of label distribution + regression is better than single label+ regression.

They deduce the facts that the fusion model has great advantages over the single model, and improve the accuracy of the model as well as the precision rate of each one.

(2) MORPH dataset

The above experiment was in the MORPH,FG-NET fusion and amplification of the dataset. In order to compare with algorithms, we alone use MORPH dataset as test set to estimate age, using MAE and CS curve to comparison and analysis.

Table 2. The MAE values of different methods

Method	DLA	SVR[2]	DIF	Reference[1]	Fusion model
MAE	4.77	4.37	3.80	3.25	2.94

The performance of various algorithms under the CS evaluation index be shown in figure 5. In terms of precise, fusion model is superior others, taking the best accuracy, especially in the range of error value 0-6. The accuracy has upward trend when the threshold of the difference is gradual increase.

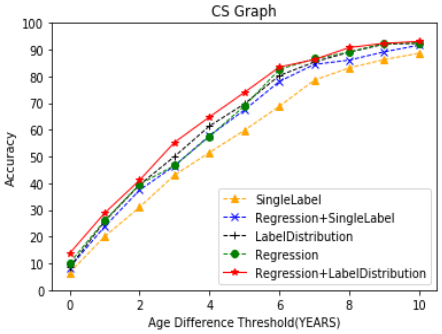

Figure 4. Multiple methods of the CS curve

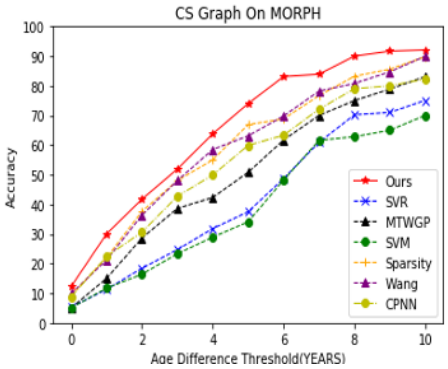

Figure 5. CS Curve on MORPH Dataset

5 Conclusions

Inspired by multi-task learning methods, for face age estimation problems, the use of both the classification and regression methods based on the shared convolutional layer parameters is better than using a single model, and the use of classification and regression fusion models only requires training. A network model greatly saves training time. In the classification model, two kinds of age-tag coding methods are used for comparison. The coding effect of the label distribution is obviously better than that of the traditional single-label coding method. The two methods are respectively integrated with the regression model, and the former experiment. The result is also better than the latter. The best experimental results were obtained using the model of label distribution + regression, indicating that the effect of the fusion model is better than that of a single model, and the encoding of label distribution is also better than the traditional single-label coding method.

References

1. Rothe R, Timofte R, Gool L V. DEX: Deep EXpectation of Apparent Age from a Single Image[C].IEEE International Conference on Computer Vision Workshop. IEEE, 2016:252-257.
2. Guo G, Fu Y, Dyer C R, et al.. Image-based human age estimation by manifold learning and locally adjusted robust regression.[J]. IEEE Transactions on Image Processing, 2008, 17(7):1178-1188.
3. Alex Krizhevsky, Ilya Sutskever, and Geoff Hinton. Imagenet classfication with deep convolutional neural networks. Inadvances in NISP,2012
4. Zhang K, Zhang Z, Li Z, et al.. Joint Face Detection and Alignment Using Multitask Cascaded Convolutional Networks[J]. IEEE Signal Processing,Letters,2016, 23(10):1499-1503.
5. Escalera S, Fabian J, Pardo P, et al..ChaLearn Looking at People 2015: Apparent Age and Cultural Event Recognition Datasets and Results[C]. IEEE International Conference on Computer Vision Workshop. IEEE, 2015:243-251.
6. Hu Z, Wen Y, Wang J, et al.. Facial Age Estimation With Age Difference[J]. IEEE Transactions on Image Processing, 2017, 26(7):3087-3097.
7. S. Ioffe and C. Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In Proceedings of The 32nd International Conference on Machine Learning, pages 448–456, 2015