

# Age Estimation Based on a Deep Model

Liming Chen\*

School of Physics & Electric Engineering, Leshan Normal University, Leshan, China

\*Corresponding author

**Abstract**—In this paper, we propose a new age classification method based on a deep model. In our method, the fine-tuned deep facial age (FTDFA) model is used to extract facial age features. Age features output from the activations of the penultimate layer are classified by the Maximum joint probability classifier(MJPCCR).Three data sets are used to validate our approach. And we also send the age features output from the activations of the last layer into the MJPC and the SVM classifier separately, and compare their results. Experiments show that, the performance of our method is superior to that of the previous methods.

**Keywords**—deep convolutional neural networks; fine-tune; face age estimation; maximum joint probability classifierstyling

## I. INTRODUCTION

The face age identification problem is an interesting subproblem of face biometrics [1], which is a very challenging problem. In recent years, CNN has been applied to face recognition because of CNN's satisfactory performance in respect of related tasks in various fields. Deep CNN models are used in, for example, face recognition [2], face alignment [3], face validation [4,5], age and gender classification [6], etc., all have achieved good results. Inspired by the above method, we use the WIKI[7] dataset to fine-tune the VGG-Face A model[8] into a new face age classification model called FTDFA model, and extract the face age features on the model , The feature is then input to the maximum joint probability classifier based on the collaborative representation (MJPCCR) for classification. We use the activations of the penultimate layer and the Last layer as local features separately, then input them to the classifier, through the experiment, we find that the local features from the activations of the penultimate layer can make the classification accuracy higher. We also send the features from the FTDFA model to the SVM classifier, then compare the results with those obtained by our method. We validate the superiority of our method on three datasets FG-NET[9], MORPH [10]and CACD[11].Experiments show that this method can improve the accuracy of face age recognition.

## II. RELATED WORK

The CNN model is an end-to-end model, the feature can be learned from raw data due to the nature of convolution and pooling calculations. The CNN model can have many layers, each layer has a lot of features. A layer of output is the next level of input. The feature extracted by the CNN model has the characteristics of semantic clustering. Since the feature detection layer learns through the training data, it avoids the explicit feature extraction and implicitly learns from the training data. Because the feature mapping surface The

neurons have the same weights, so the network can learn in parallel. The layout of the CNN is closer to the actual biological neural network. Weight sharing reduces the complexity of the network. In particular, the multi-dimensional input vector image can be directly input to the network. This feature avoids the complexity of data reconstruction during feature extraction The Through the structural reorganization and reduce the weight of the feature extraction function into the multi-layer sensor. It can directly deal with grayscale images, can be used directly to deal with image-based classification.

With the application of deep learning, people also use the CNN model to extract the features of the image and classify the face age. In recent years, X Geng et al. has proposed a number of age classification methods, mainly Label Distribution Learning (label distribution learning) age classification method [12-15]. The CNN is widely used for face age identification, such as [16-19].

## III. APPROACH

### A. Image Preprocessing

Before we extract the feature, we use the method to carry out facial images pre-processing on the input image, as shown in Figure I. We use the COC-DPM algorithm [20] for face detection, the detected face is fed to the publicly available face detector software [21] to detect five facial key points, including left / right eye center, nose tips and left / right mouth corners. Then based on all detected faces, we align the five key points of the face. After this step, we adjusted all face images based on five key points to  $256 \times 256$  pixels, then use the data augmentation method as Krizhevsky et al. described in[22]. After pretreatment, the final entry into the CNN model is a  $224 \times 224$  pixel face image.

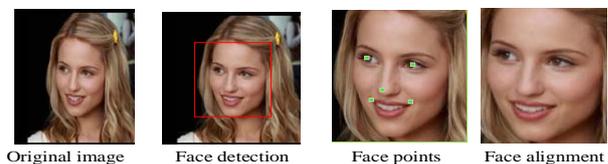


FIGURE I. IMAGE PREPROCESSING PROCESS

### B. Deep Feature Extraction Model Fine-tuning and Face Image Feature Extraction

We selected 40,000 pictures in the wiki dataset to fine-tune VGG-Face A model, these pictures were divided into 20 groups, 0-5 years old for the first group, between 6 and 100 years old, every 5 years for a group. The new model is called

FTDFA model.

The literature [8] shows the details of the FTDFA model configuration. The aligned facial images with  $224 \times 224$  pixels are input to the model. The convolutional filter is  $3 \times 3$  with stride and pad are both 1, and the max-pooling layers is  $2 \times 2$  with stride is 2. For all datasets, the final feature is 4096 dimensional. We use the activations of the penultimate layer and the Last layer as local features separately, then input them to the classifier, through the experiment, we find that the local features from the activations of the penultimate layer can make the classification accuracy higher, so in our Feature extraction model, we removed the last two fully connected layers, leaving only one of them.

### C. Maximum Joint Probability Classification Algorithm

A probabilistic collaborative representation based classifier (ProCRC) [23] was derived from [24] and [25]. Similarly, on this basis, we deduce Maximum joint probability classification based on collaborative representation (MJPCR).

Similar to [23], we can get the probability of the face image as the K class.

$$P_k = \exp\left(-\|U\sigma - U_k\sigma_k\|_2^2\right) \quad (1)$$

The classification rule is expressed as follows:

$$L(v) = \arg \max_k \{P_k\} \quad (2)$$

From the above analysis we can see that we obtain the by maximizing the joint probability based on collaborative representation. So we call the classification as Maximum joint probability classification based on collaborative representation (MJPCR).

We adopt an IRLS algorithm to compute. Similar to [30], we can finally have

$$(\sigma) = (U^T W_U U + \frac{\delta}{K} \sum_k^K \left(\frac{U_k}{U_k}\right)^T \frac{U_k}{U_k} + \omega I)^{-1} U^T W_U v \quad (3)$$

We use equation (3) to replace the  $\sigma$  in equation (1), then we can get the class corresponding  $v$  from equation (2). We alternatively update the weight matrix  $W_U$  as well as the coefficient  $\sigma$  vector, until a fixed number of iterations or convergence.

We set the parameters  $\omega$  and  $\delta$  through the 10-fold cross validation. We selected 18,000 face images in the CACD dataset, and divide them into nine groups by age, 14-20 years old for the first age group, between 21 to 55 years old, every 5 years for a group, 56-62 years old for the ninth group. Each age group has 2000 face images, which was divided into 10 groups, each group have 200 images. In the verification process, we constantly adjust the parameters and to make the

classification accuracy achieve the highest. Finally, we set the parameters of the classifier to  $\omega = 1e-2$  and  $\delta = 1e-3$ .

## IV. EXPERIMENT

We fine-tune the VGG-Face A model on the wiki dataset, and then extract the features of the three datasets FG-NE, MORPH and CACD with the FTDFA mode, and then input the feature into the MJPCR for classification. The whole Process of our method is show in Figure II. We extraction features from the activations of the penultimate layer (APLF) and the activations of the last layer (ALLF) of the FTDFA mode separately, and then input the features into the MJPCR for classification. We also compared their results. In order to verify the effectiveness of our method, We also send the features of the three datasets extracted from the FTDFA mode into the MJPCR to classify, and we compare the results of the classification.

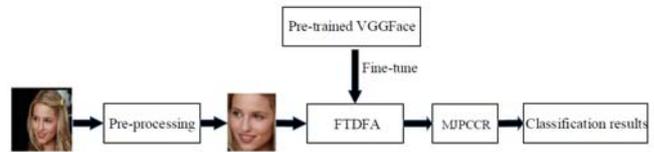


FIGURE II. THE FRAMEWORK OF PROPOSED METHOD

### A. Datasets

In this paper, four data sets are used, FG-NE dataset, MORPH dataset, CACD dataset, WIKI dataset, the WIKI datasets is used to fine-tune the model, and the other three datasets are used for experimental verification.

#### Implementation details

### B. Experiment on FG-NET Dataset

We divided FG-NET into four groups by age group, respectively 0-6 years young children, 7-17 years juvenile group, 18-40 years youth group, above 40 years middle and old age group. There are 234 face images in the training set and 40 face images in the test set in young children group, 306 face images in the training set and 60 face images in the test set in juvenile group, 252 face images in the training set and 50 face images in the test set in youth group, and 50 face images in the training set and 10 face images in the test set in middle and old age group. That is, there are total 842 face images in training set and total 160 face images in test set.

The classification accuracy is shown in Figure III.

It can be seen that when the feature is output from the same layer the maximum classification accuracy of the MJPCR classifier is 96.2% and the maximum average classification accuracy is 82.6%. the maximum classification accuracy of the SVM classifier is 90.7% and the maximum average classification accuracy is 78.0%. Therefore, the MJPCR has higher classification accuracy than the SVM classifier. From the age point of view, 0-6years group has the highest classification accuracy of 96.2%, the lowest classification accuracy of 72.2%, has the highest classification accuracy. From the output feature point of view, when the MJPCR is used, the highest classification accuracy of the

output of the activations of the penultimate layer is 96.2% and the average classification accuracy is 82.6%, the highest classification accuracy of the output of the activations of the last layer is 64.6% and the average classification accuracy is 58.6%. when the SVM classifier is used, the highest classification accuracy of the output of the activations of the penultimate layer is 90.7% and the average classification accuracy is 78.0%, the highest classification accuracy of the output of the activations of the last layer is 59.4% and the average classification accuracy is 52.5%. Therefore, in the case of the same classifier, the classification accuracy of the age features output of the activations of the penultimate layer is higher.

### C. Experiment on MORPH Dataset

We divide the MORPH datasets into 10 groups by age, between 16 to 60 years old, every 5 years for a group, above 60 years old for the tenth group. We randomly selected about one-tenth of each group as a test set, a total of 5220 images as a test set, and the remaining 47892 images as a training set.

The classification accuracy is shown in Figure III.

It can be seen that when the feature is output from the same layer, the maximum classification accuracy of the MJPCCR classifier is 98.7% and the maximum average classification accuracy is 82.7%. the maximum classification accuracy of the SVM classifier is 91.4% and the maximum average classification accuracy is 78.5%. Therefore, the MJPCCR has higher classification accuracy than the SVM classifier. From the age point of view, 16-20years group has the highest classification accuracy of 98.7%, the lowest classification accuracy of 62.5%, has the highest classification accuracy. From the output feature point of view, when the MJPCCR is used, the highest classification accuracy of the output of the activations of the penultimate layer is 98.7% and the average classification accuracy is 82.7%, the highest classification accuracy of the output of the activations of the last layer is 71.9% and the average classification accuracy is 59.2%. when the SVM classifier is used, the highest classification accuracy of the output of the activations of the penultimate layer is 91.4% and the average classification accuracy is 78.5%, the highest classification accuracy of the output of the activations of the last layer is 62.6% and the average classification accuracy is 53.5%. Therefore, in the case of the same classifier, the classification accuracy of the age features output of the activations of the penultimate layer is higher.

### D. Experiment on CACD Dataset

When we set the parameters in the classifier, we selected 18,000 face images, now we selected 36,000 face images in the remaining images, and divided them into nine groups by age. 14-20 years old for the first age group, between 21 to 55 years old, every 5 years for a group, 56-62 years old for the ninth group. Each group has 4000 face images, which randomly selected 800 as a test set, the rest for the training set. So the test set has a total of 7200 face images, training set has a total of 28800 face images.

The classification accuracy is shown in Figure III.

It can be seen that when the feature is output from the same layer the maximum classification accuracy of the MJPCCR classifier is 98.6% and the maximum average classification accuracy is 84.1%. the maximum classification accuracy of the SVM classifier is 96.6% and the maximum average classification accuracy is 79.9%. Therefore, the MJPCCR has higher classification accuracy than the SVM classifier. From the age point of view, 31-35years group has the highest classification accuracy of 98.6%, the lowest classification accuracy of 59.4%, has the highest classification accuracy. From the output feature point of view, when the MJPCCR is used, the highest classification accuracy of the output of the activations of the penultimate layer is 98.6% and the average classification accuracy is 84.1%, the highest classification accuracy of the output of the activations of the last layer is 67.4% and the average classification accuracy is 59.4%. when the SVM classifier is used, the highest classification accuracy of the output of the activations of the penultimate layer is 96.6% and the average classification accuracy is 79.9%, the highest classification accuracy of the output of the activations of the last layer is 60.7% and the average classification accuracy is 54.1%. Therefore, in the case of the same classifier, the classification accuracy of the age features output of the activations of the penultimate layer is higher.

### E. Validation Result

The classification accuracy obtained by using the three data sets in the method of this paper is shown in Figure III.

From Figure III, we can see that, when the features output layer is the same, the MJPCCR classifier is better than the SVM classifier, when the classifier is the same, the classification accuracy of the age features output of the activations of the penultimate layer is higher than that output of the activations of the last layer. Therefore, we can know that the MJPCCR classifier can get a better classification effect than the SVM classifier, the classification accuracy of the age features output of the activations of the penultimate layer is higher, the best effect can be achieved by classifying the age features of the FT DFA model's penultimate activation layer with the MJPCCR classifier.

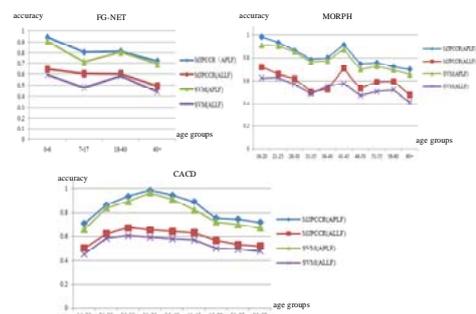


FIGURE III. THE CLASSIFICATION ACCURACY OBTAINED BY USING THE THREE DATA SETS

## V. CONCLUSION

In this paper, we propose a new method to estimate the face age, use the face age database to fine-tune a deep

convolution neural network model to obtain a new face age estimation model, and use this new model to extract the age features, and then these features are sent into a new age classification model for classification. Experiments show that the MJPCCR classifier can get a better classification effect than the SVM classifier, the classification accuracy of the age features output of the activations of the penultimate layer is higher, the best effect can be achieved by classifying the age features of the FT DFA model's penultimate activation layer with the MJPCCR classifier.

#### REFERENCES

- [1] W. Zhao, R. Chellappa, P. J. Phillips. "Face Recognition: a literature survey". *ACM Computer surveys*, 35(4):399–458, 2003
- [2] Y. sun, Y. Chen, U. Wang, and U. Tang. "Deep learning face representation by joint identification-verification". In *NIPS*, pages 1988–1996, 2014.
- [3] Y. sun, U. Wang, and U. Tang. "Deep convolutional network cascade for facial point detection". In *CVPR*, pages 3476– 3483, 2013.
- [4] Y Taigman, M Yang, M Ranzato, L Wolf. "Deep face: Closing the gap to human-level performance in face verification". In *CVPR*, pages 1701–1708, 2014.
- [5] GB Huang, H Lee, E Learned-Miller. "Learning hierarchical representations for face verification with convolutional deep belief networks". In *CVPR*, pages 2518–2525, 2012.
- [6] G. Levi and T. Hassner. "Age and gender classification using convolutional neural networks". In *CVPR*, 2015
- [7] R Rothe, R Timofte, LV Gool. "DEX:Deep expectation of apparent age from a single Image". In *ICCVW*, pages 252–257, 2015
- [8] Omkar M. parkhi, andrea Vedaldi and andrew Zisserman. "Deep Face Recognition". In *BMVC*, 2015
- [9] Face and gesture recognition, fg-net aging database. <http://www.fgnet.rsunit.com/index.php?page=parta>
- [10] K Ricanek, T Tesafave. "Morph: a longitudinal image database of normal adult age-progression". *International Conference on automatic Face & Gesture Recognition*, pages 341-345, 2006
- [11] Bor-Chun Chen, Chu-song Chen, Winston H. Hsu. "Cross-age reference coding for age-Invariant face recognition and retrieval". In *ECCV*, pages 768-783, 2014
- [12] X. Geng, K. smith-Miles, Z.-H. Zhou. "Facial age Estimation by learning from label Distributions". In *Conference on artificial Intelligence*, pages 451-456, 2010
- [13] X. Geng, Q. Wang, and Y. Xia. "Facial Age Estimation by Adaptive Label Distribution Learning". *IEEE Transactions on pattern analysis and Machine Intelligence*, 35(10): 2401-2412, 2013
- [14] X. Yang, X. Geng and D.Y. Zhou. "Sparsity Conditional Energy Label Distribution Learning for Age Estimation". In *proceedings of the International Joint Conference on artificial Intelligence*, pages 2259-2265.
- [15] Z. He, X. Li, Z. Zhang, F. Wu, X. Geng et al. "Data-Dependent label Distribution learning for age Estimation". *IEEE Transactions on Image processing*, pages 1-13, 2017
- [16] S Escalera, J Fabian, P Pardo et al. "ChaLearn Looking at People 2015: Apparent Age and Cultural Event Recognition datasets and results". In *ICCVW*, pages 243-251, 2015
- [17] X Liu, S Li, M Kan, J Zhang et al. "AgeNet Deeply Learned Repressor and Classifier for Robust Apparent Age Estimation". In *ICCVW*, pages 258-266, 2015
- [18] RC Malli, M Aygun, HK Ekenel. "Apparent age Estimation Using Ensemble of Deep learning Models". In *CVPR*, pages 714-721, 2016
- [19] FS Abousaleh, T Lim, WH Cheng, NH Yu et al. "A novel comparative deep learning framework for facial age estimation". *Eurasip Journal on Image & Video processing*, 2016:47, 2016
- [20] M Mathias, R Benenson, M Pedersoli, et al. "Face detection without bells and whistles". In *ECCV*, pages 720–735, 2014.
- [21] Y. Sun, X. Wang, and X. Tang. "Deep convolutional network cascade for facial point detection". In *CVPR*, pages 3476–3483, 2013.
- [22] A. Krizhevsky, I. Sutskever, and G. E. Hinton. "Imagenet classification with deep convolutional neural networks". *International Conference on Neural Information processing systems*, pages 1097–1105, 2012.
- [23] S Cai, L Zhang, W Zuo, X Feng. "A probabilistic Collaborative Representation based approach for pattern Classification". In *CVPR*, pages 2950-2959, 2016
- [24] J Waqas, Y Zhang, Z Lei. "Collaborative neighbor representation using l2 minimization". *Pattern Recognition letters*, 34(2):201-208, 2013
- [25] W Yang, Z Wang, C Sun. "A collaborative representation based projections method for feature extraction". *Pattern Recognition*, 48(1): 20-27, 2015