A Review on Data Dependent Label Distribution Learning for Age Estimation using CNN

Ms.Anuja Admane[1], Mr. Rahul Dhuture[2]

[1]PG Student, Department of ETC Engineering, Tulsiramji Gaikwad Patil College of Engineering, Nagpur
[2]Asst Professor, Department of ETC Engineering, Tulsiramji Gaikwad Patil College of Engineering, Nagpur

---***---

Abstract— In this paper, we endorse a new age type technique primarily based on a deep model. In our method, the great-tuned deep facial age (FTDFA) version is used to extract facial age features. Age features output from the activations of the penultimate layer are categorized by way of the Maximum joint chance classifier(MJPCCR).Three facts units are used to validate our method. And we also send the age functions output from the activations of the ultimate layer into the MJPC and the SVM classifier one at a time, and evaluate their results. Experiments show that, the overall performance of our method is advanced to that of the preceding strategies.

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

1. INTRODUCTION

The face age identity trouble is an exciting subproblem of face biometrics [1], that's a very tough hassle. In recent years, CNN has been carried out to face popularity because of CNN's great performance in respect of related responsibilities in diverse fields. Deep CNN models are used in, as an example, face reputation [2], face alignment [3], face validation [4,5], age and gender type [6], and so on., all have performed top results. Inspired via the above method, we use the WIKI[7] dataset to exceptional-music the VGG-Face A version[8] into a brand new face age class model known as FTDFA version, and extract the face age capabilities at the model , The characteristic is then input to the most joint probability classifier based totally at the collaborative illustration (MJPCCR) for type. We use the activations of the penultimate layer and the Last layer as neighborhood functions one at a time, then enter them to the classifier, through the test, we discover that the neighborhood functions from the activations of the penultimate layer could make the type accuracy higher. We additionally send the functions from the FTDFA model to the SVM classifier, then evaluate the outcomes with those obtained by means of our technique. We validate the superiority of our technique on three datasets FG-NET[9], MORPH [10]and CACD[11].Experiments show that this approach can enhance the accuracy of face age popularity.

2. RELATED WORK

The CNN version is an give up-to-cease version, the feature may be found out from raw facts because of the nature of convolution and pooling calculations. The CNN model could have many layers, every layer has lots of functions. A layer of output is the subsequent degree of input. The characteristic extracted with the aid of the CNN model has the traits of semantic clustering. Since the characteristic detection layer learns thru the schooling statistics, it avoids the express characteristic extraction and implicitly learns from the education information. Because the function mapping surface The neurons have the same weights, so the network can learn in parallel. The format of the CNN is closer to the real organic neural community. Weight sharing reduces the complexity of the network. In particular, the multi-dimensional input vector image can be without delay enter to the community. This function avoids the complexity of facts reconstruction for the duration of function extraction The Through the structural reorganization and decrease the burden of the characteristic extraction characteristic into the multi-layer sensor. It can immediately address grayscale pix, may be used at once to cope with picture-primarily based class.

With the application of deep learning, people additionally use the CNN model to extract the functions of the photograph and classify the face age. In recent years, X Geng et al. Has proposed some of age classification techniques, specially Label Distribution Learning (label distribution studying) age category approach [12-15]. The CNN is extensively used for face age identification, along with [16-19].

3. 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 ×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 ×224 pixel face image.

Original image Face detection Face points Face alignment

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 × 224 pixels are input to the model. The convolutional filter is 3 ×3 with stride and pad are both 1, and the max -pooling layers is 2 ×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(MJPCCR).

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

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

The classification rule is expressed as follows:

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

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 (MJPCCR).

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

$$(\sigma) = \left(U_U^T W U + \frac{\delta}{K}\sum_k \left(\overline{U_k}\right)^T \overline{U}_k + \omega I\right)^{-1} U^T W\, y_U \qquad (3)$$

We use equation (3) to replace the μ 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 ▯ vector, until a fixed number of iterations or convergence.

We set the parameters and 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

4. 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 MJPCCR 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 MJPCCR 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 MJPCCR to classify, and we compare the results of the classification.

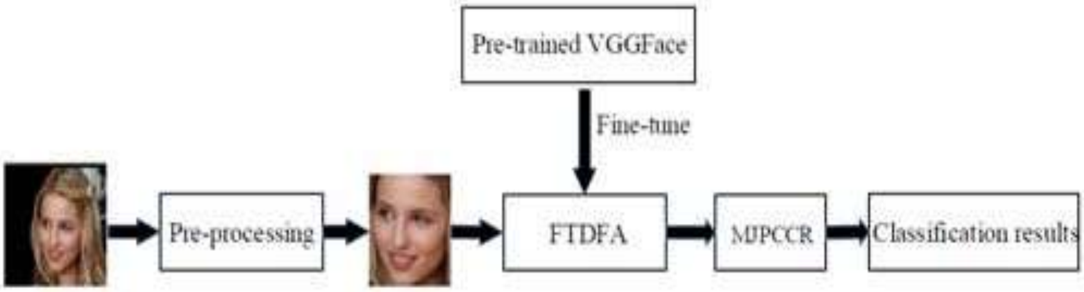

FIGURE II. THE FRAMEWORK OF PROPOSED METHOD

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 using age institution, respectively 0-6 years younger children, 7-17 years juvenile institution, 18-forty years teens institution, above 40 years center and old age group. There are 234 face images in the education set and 40 face pix inside the check set in younger children group, 306 face pics in the schooling set and 60 face photographs within the take a look at set in juvenile group, 252 face photographs inside the education set and 50 face snap shots within the test set in teenagers institution, and 50 face images in the education set and 10 face snap shots within the take a look at set in center and old age organization. That is, there are overall 842 face pix in schooling set and overall a hundred and sixty face photographs in test set.

It may be visible that after the characteristic is output from the equal layer the most category accuracy of the MJPCCR classifier is 96.2% and the maximum common class accuracy is 82.6%. The most class accuracy of the SVM classifier is 90.7% and the maximum common category accuracy is 78.0%. Therefore, the MJPCCR has higher category accuracy than the SVM classifier. From the age factor of view, 0-6 years institution has the very best category accuracy of 96.2%, the lowest type accuracy of 72.2%, has the highest type accuracy. From the output feature factor of view, whilst the MJPCCR is used, the very best classification accuracy of the Output of the activations of the penultimate layer is 96.2% and the common class accuracy is 82.6%, the very best classification accuracy of the output of the activations of the last layer is 64.6% and the common class accuracy is 58.6%. While the SVM classifier is used, the highest type accuracy of the output of the activations of the penultimate layer is 90.7% and the average category accuracy is 78.0%, the very best class accuracy of the output of the activations of the ultimate layer is 59.4% and the average class accuracy is 52.5%. Therefore, in the case of the equal classifier, the category 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 can be seen that when the feature is output from the same Layer, the most type accuracy of the MJPCCR classifier is 98.7% and the maximum common class accuracy is 82.7%. The maximum type accuracy of the SVM classifier is 91.4% and the most common classification accuracy is 78.5%. Therefore, the MJPCCR has better category accuracy than the SVM classifier. From the age point of view, 16 – 20 years institution has the very best type accuracy of 98.7%, the bottom category accuracy of 62.5%, has the very best category accuracy. From the output characteristic point of view, whilst the MJPCCR is used, the best class accuracy of the output of the activations of the penultimate layer is 98.7% and the average type accuracy is 82.7%.

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 FTDFA model's penultimate activation layer with the MJPCCR classifier. output of the activations of the ultimate layer is 71.9% and the common class accuracy is 59.2%. Whilst the SVM classifier is used, the exceptional magnificence accuracy of the output of the activations of the penultimate layer is 91.4% and the not unusual class accuracy is 78.5%, the very best kind accuracy It may be seen that when the characteristic is output from the identical layer the most type accuracy of the MJPCCR classifier is 98.6% and the maximum average classification accuracy is 84.1%. The maximum type accuracy of the SVM classifier is 96.6% and the maximum average type accuracy is 79.9%. Therefore, the MJPCCR has higher type accuracy than the SVM classifier. From the age thing of view, 31-35years agency has the highest classification accuracy of 98.6%, the bottom elegance accuracy of 59.4%, has the very nice type accuracy. From the output function point of view, at the same time as the MJPCCR is used, the best kind accuracy of the output of the activations of the penultimate layer is 98.6% and the common type accuracy is 84.1%, the very best class accuracy of the output of the activations of the remaining layer is 67.4% and the average category accuracy is 59.4%. While the SVM classifier is used, the high-quality class accuracy of the output of the activations of the penultimate layer is 96.6% and the common class accuracy is 79.9%, the highest type accuracy of the output of the activations of the very last layer is 60.7% and the average kind accuracy is 54.1%. Therefore, in the case of the equal classifier, the class accuracy of the age features output of the activations of the penultimate layer is better. 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.

5. CONCLUSION

In this paper, we firstly analyze that Ranking-based methods are implicitly learning label distribution. This result unifies two existing popular state-of-the-art age estimation methods into the DLDL framework. Second, we propose a DLDL-v2 framework which alleviates the inconsistency between training and evaluation stages via jointly learning age distribution and regressing single age with a thin and deep network architecture. The proposed approach creates new state-of-the-art results on apparent and real age estimation tasks with fewer parameters and faster speed. In addition, our DLDL-v2 is also an interpretable deep framework which employs different patterns to estimate age.

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 analvsis 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 etal. "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) 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