Mechatronics and Automation Technology
J. Xu (Ed.)

doi:10.3233/ATDE221152

Unsupervised Face Image Quality Assessment Based on Face Recognition

Lei HUO[a,1], Yanwei XIONG[a], Jian SUN[b], Yuhu NIE[a] and Wei SU[b]
[a] *Beijing Smart Chip Microelectronics Technology Company Limited Beijing, 100192, China*
[b] *State Grid Jiangsu Electric Power co.,Ltd. Electric Institute, Nanjing, China*

Abstract. Face quality evaluation can filter out low quality face image to save computational resources and improve the system performance, labeling the face image quality score by manual consume too much manpower. To solve this problem, an unsupervised face image evaluation based on face recognition is proposed. We use the face recognition model to calculate the features of faces and label the images quality score . The face recognition model is compressed by knowledge distillation method to obtain efficient quality assessment model. Experimental results show that this method can effectively evaluate the quality of face image and improve the performance of face recognition.

Keywords. Assessment of face quality, quality annotation, distillation of knowledge.

1. Introduction

Face recognition is a biometric identification technology based on facial feature information identification, similar biometric technology has fingerprint recognition, iris recognition and so on. Face recognition is easier to be accepted by users because of its non-coercive, non-contact, concurrency and other characteristics. However, in some unconstrained scenarios, such as monitoring, a large part of face images collected by the system are low quality. Face recognition of these low-quality face images is often ineffective and wasteful of computing resources. Therefore, the efficiency of the whole face system can be improved by calculating the quality score of the image and deleting the image with low score.

In the current face image quality score calculation method, the method based on image analysis mainly uses some hand-defined features. For example, Nasrollahi et al. [1] proposed the method of scoring according to attitude, brightness, resolution, etc., and finally weighted fusion to obtain the overall quality score. X.Gao et al. [2] proposed the method based on facial asymmetry. P. J. Phillips et al. [3] proposed the method based on 12 features divided into three categories (DIP features, sensor-related features and classifier related features). However, these hand-designed features are difficult to accurately and comprehensively describe the quality of face images.

[1] Corresponding Author, Lei HUO, Beijing Smart Chip Microelectronics Technology Company Limited Beijing, 100192, China; E-mail: huolei@sgchip.sgcc.com.cn

Another method that has been heavily used and studied is the deep learning method. The method based on supervision requires a large number of annotated face image data sets, and it is very difficult to obtain such data sets, and the manually annotated score is difficult to accurately describe the image quality and ensure the consistency of the score. The method based on unsupervised can avoid this problem, a number of studies generate facial image quality score using auxiliary model automatically to train face image quality evaluation model. For example, Hernandez-Ortega et al. [4] proposed a method that takes the Euclidean distance between the image and the face recognition features of the in-class image as the mass fraction. Terhorst et al. [5] proposed the method that take the difference in the output characteristics of different face recognition model as the quality score.

In this paper, we propose a new method to evaluate face quality score based on face recognition. The main improvement directions are as follows:

1) A new generation method of face image quality score label is proposed, which can generate face quality score unsupervised on the face recognition data set. This method integrates image visual quality and face recognition feature quality. CNN low-level features are used to represent image visual quality and high-level features are used to calculate face recognition feature quality. When calculating the feature quality of face recognition, the fusion considers the feature quality of single image and the similarity of similar features.

2) We train the efficient face recognition model by knowledge distilling. The lightly face image as a quality model of weight training.

The training of face quality assessment model needs face image quality score as label. The main work of this paper is in the process of generating face image quality score.

2. Face Image Quality Score Generation

Under normal circumstances, the image quality refers to the image visual quality, such as brightness, resolution, resolution, etc., but these indicators cannot complete describe the image quality for face recognition system, so in the generated image quality score, we joined the quality score calculated by the face recognition model, and use the low-level features of face recognition model to replace the image visual quality.

2.1. Image Visual Quality Score Calculation

Deep neural network uses multi-layer neurons to process the original image to obtain the final desired features. In general, features obtained by high-level neurons are used to describe the semantic information of the image, while features obtained by low-level neurons are mainly used to describe the visual features of the image. So wo use the low-level features of face recognition network to describe the visual feature Fv of images. S_{vi} is defined as the quality fraction of image I, and its calculation method is as follows:

$$S_{vi} = 1 - \frac{d_i}{d_{max} - d_{min}} \tag{1}$$

Where d_i represents the Euclidean distance between the visual feature of image i and the visual feature center point Fv_m of all images in category M to which i belongs. d_{max} represents the maximum value of the Euclidean distance between all images in category m and Fv_m, and d_{min} represents the minimum value of Euclidean distance between all images in category m and Fv_m. Fv_m is calculated as follows:

$$Fv_m = \frac{\sum_{i=1}^{N} Fv_i}{N},\ i \in m \tag{2}$$

2.2. Face Recognition Feature Quality Score Calculation

This paper evaluates the quality of face recognition features from two aspects: the quality of single image and the similarity of face features with the same ID.

1) In this paper, the face image quality evaluation method proposed in SER-FIQ[5] is used to calculate the face recognition feature quality fraction S_{ri} of a single image. A trained face recognition model with Dropout layer is used to extract the features of face images. By using different Dropout modes, m embedding x_s of random subnetworks is generated. The random embedding of the same image constitutes a set $X(I) = \{x_s\}, s \in \{1,2, \dots \dots, m\}$. Then, S_{ri} can be calculated as follows:

$$S_{ri} = 2\sigma(-\frac{2}{m^2}\sum_{p<q} d(x_p, x_q)) \tag{3}$$

Among them, the $\sigma(.)$ Represents the sigmoid function, and $d(x_p, x_q)$ represents the Euclidean distance between two random embeddings. The greater the random embedding variation between the same image, the lower the robustness of the representation feature, the worse the quality of the image. On the contrary, the smaller the change, the better the robustness of the representative feature, the higher the image quality.

2) As the input image of face recognition model, the higher the similarity of face features between the same ID, the easier it is to correctly identify the face, the higher the image quality. In this paper, the similarity between the features of all face images with the same ID is used to calculate the quality fraction S_{ci}, and the calculation method is as follows:

$$S_{ci} = \frac{\sum_{j=1}^{N} (\cos(Fc_i, Fc_j) + 1)/2}{N},\quad Id_i = Id_j \tag{4}$$

Where, Fc_i represents the face recognition features of image I, $\cos(.)$ Represents cosine similarity, image i and image j are different face images with the same ID. The average cosine similarity of different face image features in the same ID is taken as its quality fraction.

The three face quality scores are weighted and fused to generate the final face image quality score S, which is calculated as follows:

$$S_i = \propto * S_{vi} + \beta * S_{ri} + \gamma * S_{ci} \tag{5}$$

Finally, the generated face quality scores were used as labels to train the face quality assessment model.

3. Lightweight Face Quality Assessment Model

In order to use the face recognition model to do knowledge transfer, most face quality assessment model use the similar structure with the face recognition network. For guaranteeing the accuracy of face recognition, currently face recognition model used relatively large amount of calculation usually. Add a network with the same structure of face quality evaluation model to the face recognition system, often unfriendly to mobile devices.

Knowledge distillation[6] adopts the teacher-student model: The complex and large model is taken as the Teacher, and the structure of the Student model is relatively simple. The Teacher is used to assist the training of the Student model. The Teacher has strong learning ability, and the knowledge learned by the Teacher can be transferred to the Student model with relatively weak learning ability. In order to enhance the generalization ability of Student model.

We use face recognition model to be the Teacher model, and a model with small structure is defined as Student model. The output of face recognition feature of Teacher model is F_t, and the output of face ID classification is C_t. The output of face recognition feature of Student model is F_s, and the output of face ID classification is C_s. Face ID classification label is C_l then Loss in Student model training is:

$$Loss = \alpha * MSE(F_s, F_t) + \beta * CE(C_s, C_t) + \gamma * CE(C_s, C_l) \quad (6)$$

Where, $MSE(F_s, F_t)$ Represents the mean square error loss function, $CE(.)$ Is the cross entropy loss function. After obtaining Student model according to the above loss training, the same Backbone structure as Student model is used to carry out knowledge transfer when training face quality assessment model.

4. Experiment

4.1. Setting

Dataset: MS-Celeb-1M [7] dataset was used to train face recognition and face quality assessment models. In the test phase, LFW [8] data set, Adience [9] data set and IJB-C [10] data set were used.

Face recognition model: ResNet101[11] model trained on MS-Celeb-1M dataset is Teacher model, and the face recognition features generated are 256 dimensional vectors. ResNet34 model was used as Student model, and knowledge distillation training was also conducted on ms-celeb-1m dataset.

Implementation details: All models were built and trained using the PyTorch framework in Ubuntu and accelerated using 4 NVIDIA Titan XP. The input size of face recognition model and quality assessment model is 112x112. Adam was used in the optimization method, and the initial learning rate was 0.001, and the learning rate was multiplied by 0.05 for every 10 epochs. Face image quality fraction weight α, β, γ were set to 0.2,0.4,0.4 respectively. The weights of Loss α, β and γ for knowledge distillation training were set to 0.4,0.2 and 0.4 respectively.

4.2. Result

The purpose of face image quality evaluation is mainly to optimize the effect of face recognition, improve its performance, so the evaluation of the face quality evaluation network after the impact on face recognition system. The effects of different face quality assessment models on the FNMR (False non-match rate) index of face recognition system were compared. FNMR represents the proportion of the samples rejected by the quality assessment model that can be correctly recognized to the rejected samples. Figure 1 shows FNMR curves of FaceQnet, Ser-FIQ and the proposed method on LFW, Adience and IJB-C data sets. As can be seen from the figure, FNMR values of the proposed method are slightly lower than those of FaceQNet and SER-FIQ methods at most rejection rates in the three data sets.

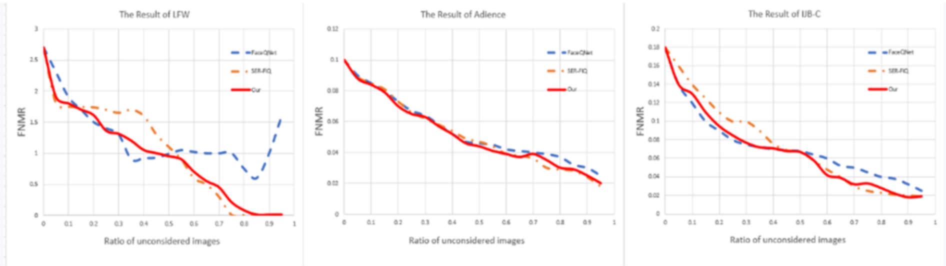

Figure 1. FNMR curces

Since knowledge distillation is used for network compression in this paper, the computation and parameter number of the three methods are compared. Table 1 lists the computation amount and parameter number of FaceQNet, SER-FIQ and the method in this paper. It can be seen that the computation amount and parameter number of the method in this paper are both smaller than those of the other two methods.

Table 1. Params and Flops

Model	**Params(M)**	**Flops(G)**
FaceQNet	44.5	2.02
SER-FIQ	44.5	2.21
Our	21.8	0.97

5. Conclusion

An unsupervised face image quality score calculation method based on face recognition is proposed. Face recognition model and its data set are used to generate face image quality score automatically. The score comprehensively considers the visual quality and face recognition feature quality of face image, and constructs a new lightweight face quality evaluation model through knowledge distillation. The experimental results show that the unsupervised face image quality evaluation method based on face recognition is slightly better than the comparison algorithm under the lower computation and parameter number, which is helpful to improve the performance of general face recognition system.

Acknowledgments

This work is supported by the the science and technology program of State grid Corporation of China (5700-202141451A-0-0-00), which is "Embedded AI Multi-level Interconnected Heterogeneous Multi-core System-on-Chip (SoC) Architecture Research and Chip Development".

Reference

[1] Nasrollahi K, Moeslund T B. Extracting a good quality frontal face image from a low-resolution video sequence[J]. IEEE transactions on circuits and systems for video technology, 2011, 21(10): 1353-1362.

[2] Gao X, Li S Z, Liu R, et al. Standardization of face image sample quality[C]//International Conference on Biometrics. Springer, Berlin, Heidelberg, 2007, pp. 242-251.
[3] Phillips P J, Beveridge J R, Bolme D S, et al. On the existence of face quality measures[C]//2013 IEEE Sixth International Conference on Biometrics: Theory, Applications and Systems (BTAS). IEEE, 2013, pp.1-8.
[4] Hernandez-Ortega J, Galbally J, Fierrez J, et al. Faceqnet: Quality assessment for face recognition based on deep learning[C]//2019 International Conference on Biometrics (ICB). IEEE, 2019, pp.1-8.
[5] Terhorst P, Kolf J N, Damer N, et al. SER-FIQ: Unsupervised estimation of face image quality based on stochastic embedding robustness[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2020, pp. 5651-5660.
[6] Hinton G, Vinyals O, Dean J. Distilling the knowledge in a neural network [J]. arXiv preprint arXiv:1503.02531, 2015, 2.
[7] Deng J, Guo J, Xue N, et al. Arcface: Additive angular margin loss for deep face recognition[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2019, pp.4690-4699.
[8] Huang G B, Mattar M, Berg T, et al. Labeled faces in the wild: A database forstudying face recognition in unconstrained environments[C]//Workshop on faces in'Real-Life'Images: detection, alignment, and recognition, 2008.
[9] Eidinger E, Enbar R, Hassner T. Age and gender estimation of unfiltered faces[J]. IEEE Transactions on information forensics and security, 2014, 9(12): 2170-2179.
[10] Maze B, Adams J, Duncan J A, et al. Iarpa janus benchmark-c: Face dataset and protocol[C]//2018 international conference on biometrics (ICB). IEEE, 2018, pp. 158-165.
[11] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016, pp. 770-778.