

Cross domain meta-network for sketch face recognition

Yuying Shao¹, Lin Cao^{1,*}, Changwu Chen¹, and Kangning Du¹

¹ School of Information and Communication Engineering, Beijing Information Science and Technology University, China

Abstract. Because of the large modal difference between sketch image and optical image, and the problem that traditional deep learning methods are easy to overfit in the case of a small amount of training data, the Cross Domain Meta-Network for sketch face recognition method is proposed. This method first designs a meta-learning training strategy to solve the small sample problem, and then proposes entropy average loss and cross domain adaptive loss to reduce the modal difference between the sketch domain and the optical domain. The experimental results on UoM-SGFS and PRIP-VSGC sketch face data sets show that this method and other sketch face recognition methods.

1 Introduction

Sketch face recognition[1] is the process of matching face photos with sketch images. Because of its important role in criminal investigations, it has received great attention in recent years[2].

To solve the problem of modal differences, the algorithms currently designed for sketch face recognition can be divided into traditional sketch face recognition methods and deep learning-based methods. The traditional sketch face recognition methods can be roughly divided into three categories[3]: feature-based methods, the method based on synthesis and the method based on common subspace projection. Feature-based methods aim to represent face images with local feature descriptors. However, this type of method usually results in the loss of details in the face image. The synthesis-based method solves the problem of face photos (photos) by converting photos (sketches) into sketches (photos), and then using the facial recognizer to match the synthesized sketches (photos) with the original sketches (photos). However sketching face synthesis[4] is more challenging. The method based on subspace projection aims to project facial images of different modalities into a common subspace to reduce the influence of modal differences on recognition performance. However, this type of method may lose important information in the original image.

To solve the above problems, a Cross Domain Meta-Network for sketch face recognition is proposed here. In addition, the entropy mean loss and cross domain adaptive loss are proposed to reduce the modal difference between the sketch domain and the optical domain and improve the recognition ability of the model.

* Corresponding author: charlin26@163.com

2 Cross domain meta-network

In this section, the cross domain adaptive entropy element network is introduced in detail. Figure 1 shows the training process of a single training episode and a single test episode.

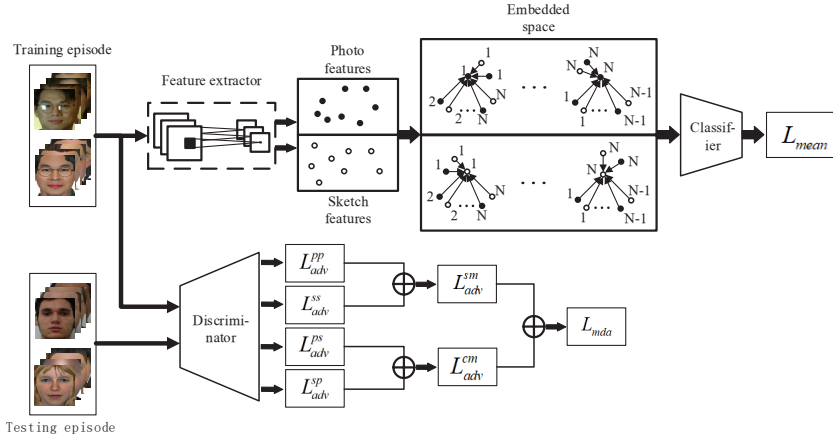


Fig. 1. The training process of a single training episode and a single test episode. First, the feature extractor extracts the features in the training episode, and then learns the metric relationship between the two in the embedding space, and finally classifies the image through the meta-classifier and calculates the entropy mean loss.

2.1 Mean entropy loss

Randomly sample $K < N$ classes from the training set $D_{\text{train}} = \{(I_1^p, y_1), (I_1^s, y_1), \dots, (I_N^p, y_N), (I_N^s, y_N)\}$ with N classes to form a training episode, where I^p and I^s are photos and sketches respectively, and y is the corresponding label. In each training episode, all images constitute the query set $Q = \{(I_1^p, y_1), (I_1^s, y_1), \dots, (I_K^p, y_K), (I_K^s, y_K)\}$, all the photos constitute the photo support set $S_p = \{(I_1^p, y_1), \dots, (I_K^p, y_K)\}$ and all the sketches constitute the sketch support set $S_s = \{(I_1^s, y_1), \dots, (I_K^s, y_K)\}$. In the test set D_{train} , randomly sampled K photos to form a photo test episode E_p and K sketches to form a sketch episode E_s . Use $f(\cdot)$ to represent the embedding function implemented by the feature extractor. Given a query set image I_c , the corresponding feature vector is $f(I_c)$. The Euclidean distances between it and the photo $I_i^p (i \in 1, \dots, K)$ in the photo support set S_p and the sketch $I_i^s (i \in 1, \dots, K)$ in the sketch support set S_s are:

$$d_{ci}^p = \|f(I_c) - f(I_i^p)\|_2, \quad d_{ci}^s = \|f(I_c) - f(I_i^s)\|_2 \quad (1)$$

where $\|\cdot\|_2$ is the Euclidean distance. Therefore, the probabilities in the support set are:

$$p(y = y_i | I_c) = \exp\left(\frac{1}{2}(d_{ci}^p + d_{ci}^s)\right) / \sum_{j=1}^K \exp\left(\frac{1}{2}(d_{ci}^p + d_{ci}^s)\right) \quad (2)$$

The mean entropy loss is obtained:

$$L_{\text{mean}} = \frac{1}{|Q|} \sum_{(I_j, y_j) \in Q} -\ln \left[M_{c_j} + \ln \sum_{j=1}^K \exp\left(-\frac{1}{2}(d_{ci}^p + d_{ci}^s)\right) \right] \quad (3)$$

2.2 Cross domain adaptive loss

Aiming at the data offset between the training data and the test data, a co-modal adversarial loss is proposed:

$$\begin{aligned} L_{adv}^{cm} = & E[\log D_{dom}(S_p) + \log(1 - D_{dom}(E_p))] + \\ & E[\log D_{dom}(E_p) + \log(1 - D_{dom}(S_p))] + \\ & E[\log D_{dom}(S_s) + \log(1 - D_{dom}(E_s))] + \\ & E[\log D_{dom}(E_s) + \log(1 - D_{dom}(S_s))] \end{aligned} \quad (4)$$

where $D_{dom}(\cdot)$ is the discriminator. Aiming at the modal difference between the sketch and the photo, a cross modal adversarial loss is proposed:

$$\begin{aligned} L_{adv}^{cm} = & E[\log D_{dom}(S_p) + \log(1 - D_{dom}(E_s))] + \\ & E[\log D_{dom}(E_s) + \log(1 - D_{dom}(S_p))] + \\ & E[\log D_{dom}(S_s) + \log(1 - D_{dom}(E_p))] + \\ & E[\log D_{dom}(E_p) + \log(1 - D_{dom}(S_s))] \end{aligned} \quad (5)$$

Combining formula (4) and formula (5), the final cross domain adaptive loss is obtained:

$$L_{mda} = L_{adv}^{sm} + L_{adv}^{cm} \quad (6)$$

The optimization goal of this article is:

$$\min L = L_{\text{mean}} + \alpha L_{mda} \quad (7)$$

where α is the balance hyperparameter between the entropy mean loss and the cross domain adaptive loss.

3 Experiments

3.1 Implementation details

This paper uses UoM-SGFS and PRIP-VSGC datasets to evaluate the effectiveness of this method. This article sets up two data sets based on the above two data sets. In the first data set (S1), 450 pairs of sketch-photos as the training set, and 150 pairs of sketch-photos are used as the test set. In the second data set (S2), the data set setting is the same as S1. This paper uses the MTCNN method to locate and align the images. Before the training process, all images are cropped to a standard size of 256×256 . In the training process, the feature extractor uses the resnet-18 network pre-trained by the ImageNet dataset. The Adam optimizer with parameters $\beta_1 = 0.5$ and $\beta_2 = 0.999$ are used to optimize the entire network model, and the initial learning rate is set to 0.0001.

3.2 Comparative experiments

In order to prove the effectiveness of the method in this paper, we compare with other advanced sketch face recognition methods, including LGMS[5], D-RS[6], D-RS+CBR[7], DEEPS[8], JAN[9] and CDAN[10]. The Table 1 and Fig. 2~3 show the experimental results of these methods and the method in this paper. It can be found that the method in this paper is better than these methods.

Table 1. Comparative experiment results on the S1 and S2 datasets.

Methods	S1			S2		
	Rank-1	Rank-10	Rank-50	Rank-1	Rank-10	Rank-50
LGMS	21.87	51.20	72.40	40.80	70.80	86.40
D-RS	22.13	49.33	69.87	42.93	75.87	90.13
D-RS+CBR	25.87	56.00	76.27	43.47	73.60	86.93
DEEPS	31.60	66.13	86.00	52.17	82.67	94.00
JAN	53.87	86.53	94.00	61.87	91.07	97.33
CDAN	57.07	88.53	95.07	62.00	91.87	97.47
Proposed method	65.63	93.15	98.33	74.33	95.19	99.33

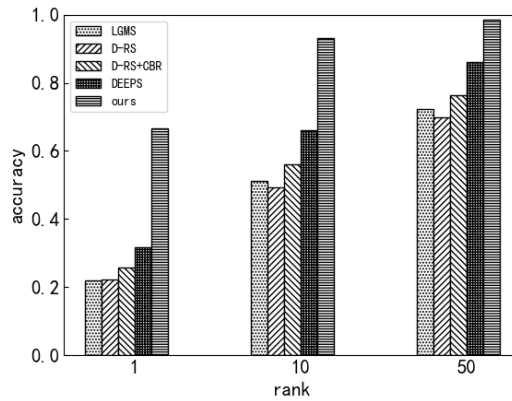


Fig. 2. S1 dataset comparison experiment results.

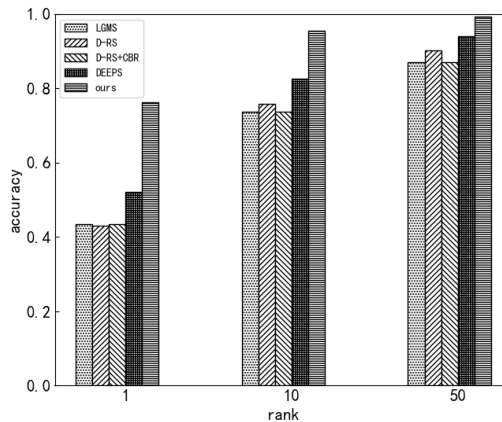


Fig. 3. S2 dataset comparison experiment results.

4 Conclusion

Aiming at the problem of the small amount of existing sketch face data and the large modal gap between sketch image and optical image, this paper proposes a sketch face recognition method based on a Cross Domain Meta-Network. This method trains the network by designing a meta-learning method based on the distance measurement and improves the learning level from data to the task, thereby improving the generalization ability of the network. In addition, by introducing average loss and cross domain adaptive loss into the cross domain adaptive entropy element network, the modal difference between the sketch domain and the optical domain is reduced. Experiments on the UoM-SGFS database and PRIP-VSGC database show that this method can effectively improve the effect of sketch face recognition.

References

1. Uhl R G J, Lobo N D V, Kwon Y H. Recognizing a facial image from a police sketch.(IEEE, 1994)
2. Peng C , Wang N , Gao X , et al. Face Recognition from Crossple Stylistic Sketches: Scenarios, Datasets, and Evaluation. (2016)
3. Wan W , Gao Y , Lee H J . Transfer deep feature learning for face sketch recognition. (Neural Computing and Applications, 2019)
4. L. Chang, M. Zhou, Y. Han and X. Deng, "Face Sketch Synthesis via Sparse Representation," (2010 20th International Conference on Pattern Recognition, Istanbul. 2010)
5. C. Galea and R. A. Farrugia, "Matching Software-Generated Sketches to Face Photographs With a Very Deep CNN, Morphed Faces, and Transfer Learning," in IEEE Transactions on Information Forensics and Security, vol. **13**.(2018)
6. C. Galea and R. A. Farrugia, "Face photo-sketch recognition using local and global texture descriptors" .(EUSIPCO, 2016)
7. B. F. Klare and A. K. Jain, "Heterogeneous Face Recognition Using Kernel Prototype Similarities," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. **35**. (2013)
8. Klum S J, Han H, Klare B F and Jain A K, The FaceSketchID system: matching facial composites to mugshots.(IEEE Transactions on Information Forensics and Security, 2014)
9. Long, Mingsheng & Wang, Jianmin & Jordan, Michael. Deep Transfer Learning with Joint Adaptation Networks. (2016).
10. Long M, Cao Z, Wang J, et al. Conditional adversarial domain adaptation. (2018)