TCN: TRANSFERABLE COUPLED NETWORK FOR CROSS-RESOLUTION FACE RECOGNITION*

Juan Zha *Hongyang Chao*†

Sun Yat-sen University, Guangzhou, P.R. China
The Key Laboratory of Machine Intelligence and Advanced Computing, Ministry of Education
chaj3@mail2.sysu.edu.cn, isschhy@mail.sysu.edu.cn

ABSTRACT

Cross-resolution face recognition (CRFR) aims to learn the matching of a low-resolution (LR) probe image with a database of high-resolution (HR) gallery images. Existing methods including super resolution and projection-based algorithms are not recognition-oriented and computationally expensive, or ignore the inter-class associations across resolutions. To address the issues, we propose a novel end-to-end Transferable Coupled Network (TCN) for CRFR. Specifically, the TCN consists of two networks for the HR and LR domains, respectively. To reduce the resolution mismatch, a transferrable triple loss (TTL) is introduced to pull together cross-resolution positive pairs (intra-class) and also enforce margins towards negative ones (inter-class) from both domains. Besides, to keep stability and faster convergence, a novel online triplet selection method is proposed. Empirically, the proposed TCN model consistently outperforms the state-of-the-art methods among various low resolutions and architectures on public LFW and SCFace benchmarks.

Index Terms— Low resolution, face recognition, domain adaptation, transferable triple loss

1. INTRODUCTION

Face recognition (FR) area has witnessed abundant achievements under various challenging scenarios over the past decades [1, 2, 3, 4, 5]. However, existing methods often assume that the region of the face images is large enough and contains sufficient detail information, which ignore resolution variations in practical. For example, due to the prohibitive costs of installing high-definition cameras all around, surveillance and monitoring systems usually rely on cameras of very limited definitions. As such, the face region can be extremely small, thereby resulting in errors when matching against high resolution images, e.g., profile images on social media or mugshot images captured by law enforcement. Therefore, cross-resolution (also mentioned as "low resolution" in some works) face recognition (CRFR), which aims to improve the learning of matching a low-resolution (LR) probe image with a database of high-resolution (HR) gallery images, has become a promising direction.

Empirical studies [6, 7, 8] have demonstrated the dramatically degraded performances of the state-of-the-art face recognition models when there exists large resolution gap. In the literature, existing methods can be generally divided into two categories. One intuitive method is to reconstruct the HR probe image given the LR input by super resolution (SR) algorithms [8, 9, 10, 11, 12]. Although the missing information can be recovered to obtain satisfactory HR images, SR-based methods could be computationally costly due to their non-end-to-end way. Besides, these methods still cannot achieve satisfactory results since they are not optimized for recognition purposes. Another line of work tries to project the HR-LR image pairs into a common feature space [13, 14, 15, 16, 17, 18], where the distance between them is optimized to be minimized. Mundununi et al. [16] propose a multi-dimensional scaling to learn a shared transformation matrix for solving the resolution variations. Lu et al. [17] propose deep coupled ResNet (DCR), where the trunck network is trained by face images of significantly different resolutions and two branch networks, trained by HR and targeted LR images, work as resolution-specific coupled mappings. Unfortunately, these methods only consider learning the intra-class mappings between the HR-LR pairs while not investigating the inter-class associations across resolutions.

To address the issue, we propose an end-to-end Transferable Coupled Network (TCN) for CRFR problem. Typically, we regard this task as a novel domain adaptation problem, where each resolution refers to a domain. The goal lies in leveraging distilled knowledge learned from the HR (source) domain to improve the matching with a LR (target) domain that lacks sufficient image details for recognition guarantees. Inspired by the triple loss, which has been successfully applied in many CV applications [1, 19], we propose a novel transferable triple loss (TTL) to effectively reduce the resolution gap. The TTL depends on pulling together positive pairs (intra-class) as the pivots between domains to push away

†Corresponding author: Hongyang Chao

*This work is partially supported by NSF of China under Grant 61672548, U1611461, 61173081, and the Guangzhou Science and Technology Program, China, under Grant 201510010165.

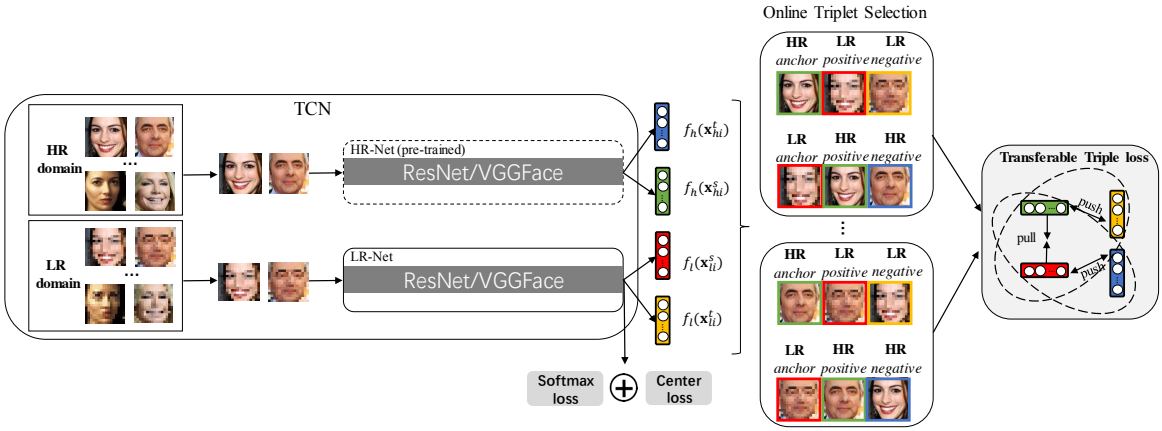

Fig. 1. The architecture of the proposed Transferable Coupled Network (TCN) model. (*Best viewed in color.*)

negative pairs (inter-class) from two different types of cross-resolution triplets simultaneously. Specifically, the TCN consists of two networks for both domains, where the HR-net that is pre-trained on HR images and fixed acts as a teacher to guide the learning of the LR-net. The LR-net is jointly trained by the proposed TTL, softmax loss and center loss [2] such that the feature representations can be both discriminative and resolution-invariant. For stability and faster convergence, we also introduce a novel cross-resolution triplet selection method to online select hard triplets within a mini-batch. We evaluate the proposed TCN model among various low resolution settings including 8x8, 12x12, 16x16 and 20x20, and two architectures VGGFace [20] and ResNet [2] on LFW and SCFace datasets. Our model outperforms SR-based methods, VDSR [10] and DRRN [9] by 8.32% and 8.12% on average on LFW dataset, respectively. Compared with projection-based methods, the proposed TCN model outperforms the best baseline DCR [17] by 1.48% and 1.88% on average on LFW and SCFace datasets, respectively.

2. METHODOLOGY

2.1. Problem Definition and Notations

Given a set of labeled HR training data $\mathbf{X}_h = \{\mathbf{x}_h^i, y_h^i\}_{i=1}^N$ from a HR domain D_h, where $\mathbf{y}_h^i$ is the class label of the i-th HR face image $\mathbf{x}_h^i$, we down-sample each HR image $\mathbf{x}_h^i$ to the targeted LR size and then up-sample it to the same size as the HR one by bicubic interpolation. As such, we can construct a set of labeled LR training data $\mathbf{X}_l = \{\mathbf{x}_l^j, y_l^j\}_{j=1}^N$ as a targeted LR domain D_l. For testing, we regard the original testing set as *gallery images* and obtain the targeted LR testing set in the same way as *probe images*. The goal of the CRFR is to match a LR probe image to the database of HR gallery images.

2.2. Overview

We propose Transferable Coupled Network (TCN), an end-to-end architecture as shown in Figure 1, to capture deep face representations that are both discriminative and resolution-invariant for cross-resolution face recognition. TCN has three key components: 1) two parallel deep convolutional neural networks (CNN), e.g., VGGFace [20] or ResNet [2], for learning deep resolution-specific representations of the HR and LR domains, respectively. The two parallel CNNs have same structure but different configurations. The HR-net is pre-trained on HR face images and fixed during the training process, which serves as a teacher to provide distilled knowledge for the LR-net, while only the LR-net needs to be learned with the aim of reducing the resolution gap across domains; 2) a novel transferable triple loss for pulling together similar cross-resolution pairs and pushing away dissimilar cross-resolution pairs among two different types of triplets; 3) a novel cross-resolution triplet selection module.

2.3. Preliminaries: Triple Loss

The triple loss proposed in FaceNet [1] for face recognition, tries to enforce a margin between each pair of faces from one person to all other faces. The loss aims to ensure that an image $\mathbf{x}_i^a$ (*anchor*) of a specific person is closer to all other images $\mathbf{x}_i^p$ (*positive*) of the same person than it is to any image $\mathbf{x}_i^n$ (*negative*) of any other person. Thus, for each triplet $(\mathbf{x}_i^a, \mathbf{x}_i^p, \mathbf{x}_i^n), i = 1, 2, ..., N$, we want,

$$\|f(\mathbf{x}_i^a) - f(\mathbf{x}_i^p)\|_2^2 + \alpha < \|f(\mathbf{x}_i^a) - f(\mathbf{x}_i^n)\|_2^2,$$

where α is a margin that is enforced between positive and negative pairs. Then the triple loss is defined as:

$$L_t = \sum_{i=1}^{N} [\|f(\mathbf{x}_i^a) - f(\mathbf{x}_i^p)\|_2^2 - \|f(\mathbf{x}_i^a) - f(\mathbf{x}_i^n)\|_2^2 + \alpha]_+$$

2.4. Transferable Triple Loss

Most of previous methods for cross-resolution face recognition focus on learning a transformation to minimize the intra-class distance while ignoring the inter-class distance across resolutions. Thus, we propose a novel transferable triple loss (TTL), which considers both of them and allows the faces for one identity with different resolutions stay on a manifold, and meanwhile, enforce the distance of other identities with different resolutions. Mathematically, we

parameterize the HR-net and LR-net by $f_h(\mathbf{x}_h) \in \mathbb{R}^d$ and $f_l(\mathbf{x}_l) \in \mathbb{R}^d$, which embeds a HR image $\mathbf{x}_h \in D_h$ and a LR image $\mathbf{x}_l \in D_l$ into a d-dimensional feature space, respectively. Besides, the feature representations are constrained to live on the d-dimensional hypersphere, i.e., $\|f_h(\mathbf{x}_h)\|_2 = 1$, $\|f_l(\mathbf{x}_l)\|_2 = 1$. The triplets are distributed across resolutions. Among the cross-resolution triplets, two different categories can be derived in term of the anchor's resolution. One type is named *HLL*-triplets $\mathcal{T}_h = \{t_{hi}\}_{i=1}^{N_h}$, where the anchor locates in the HR domain. Each triplet $t_{hi} = (\mathbf{x}_{hi}^a, \mathbf{x}_{li}^p, \mathbf{x}_{li}^n)$ behaves as (*HR-anchor*, *LR-positive*, *LR-negative*). We ensure that a HR image $\mathbf{x}_{hi}^a$ of a specific person is closer to all other LR images $\mathbf{x}_{li}^p$ of the same person than it is to any LR image $\mathbf{x}_{li}^n$ of any other person. Thus, $\forall(\mathbf{x}_{hi}^a, \mathbf{x}_{li}^p, \mathbf{x}_{li}^n) \in \mathcal{T}_h$, we want,

$$\|f_h(\mathbf{x}_{hi}^a) - f_l(\mathbf{x}_{li}^p)\|_2^2 + \beta < \|f_h(\mathbf{x}_{hi}^a) - f_l(\mathbf{x}_{li}^n)\|_2^2 \quad (1)$$

Similarly, the other type is named *LHH*-triplets $\mathcal{T}_l = \{t_{lj}\}_{j=1}^{N_l}$, where the anchor lies in the targeted LR domain. Each triplet $t_{lj} = (\mathbf{x}_{lj}^a, \mathbf{x}_{hj}^p, \mathbf{x}_{hj}^n)$ is in the form of (*LR-anchor*, *HR-positive*, *HR-negative*). We also wish a LR image $\mathbf{x}_{lj}^a$ of a specific person is closer to all other HR images $\mathbf{x}_{hj}^p$ of the same person than it is to any HR image $\mathbf{x}_{hj}^n$ of any other person. Thus, $\forall(\mathbf{x}_{lj}^a, \mathbf{x}_{hj}^p, \mathbf{x}_{hj}^n) \in \mathcal{T}_l$, we want,

$$\|f_l(\mathbf{x}_{lj}^a) - f_h(\mathbf{x}_{hj}^p)\|_2^2 + \gamma < \|f_l(\mathbf{x}_{lj}^a) - f_h(\mathbf{x}_{hj}^n)\|_2^2 \quad (2)$$

As we can see, both types of triplets can be combined together into consideration. In this way, an anchor-positive pair (same identity with different resolutions), which enforces distances of both HR and LR negative ones (another identity) simultaneously, can act as a bridge to allow fully interactions between the HR and LR domains. Given a set of quads $\mathcal{Q} = \{q_i\}_{i=1}^{N_q}$, where each quad $q_i = (\mathbf{x}_{hi}^s, \mathbf{x}_{li}^s, \mathbf{x}_{hi}^t, \mathbf{x}_{li}^t)$ consists of the images of two different identities with their HR and LR forms. The quad can be decomposed into the two types of triplets aforementioned, i.e., $(\mathbf{x}_{hi}^{sa}, \mathbf{x}_{li}^{sp}, \mathbf{x}_{li}^{tn})$ and $(\mathbf{x}_{li}^{sa}, \mathbf{x}_{hi}^{sp}, \mathbf{x}_{hi}^{tn})$. Thus, the proposed transferable triple loss is:

$$\begin{aligned} L_{ttl} = \sum_{i=1}^{N_q} & [\|f_h(\mathbf{x}_{hi}^{sa}) - f_l(\mathbf{x}_{li}^{sp})\|_2^2 - \|f_h(\mathbf{x}_{hi}^{sa}) - f_l(\mathbf{x}_{li}^{tn})\|_2^2 + \beta]_+ \\ & + [\|f_l(\mathbf{x}_{li}^{sa}) - f_h(\mathbf{x}_{hi}^{sp})\|_2^2 - \|f_l(\mathbf{x}_{li}^{sa}) - f_h(\mathbf{x}_{hi}^{tn})\|_2^2 + \gamma]_+. \end{aligned}$$

2.5. Cross-resolution Triplet Selection

Choosing suitable cross-resolution triplets is crucial to achieve fast convergence and superior performance. Thus, we propose an online cross-resolution triplet selection method, which selects all anchor-positive pairs (same identity with different resolutions) while hard negatives from within a HR-LR mini-batch pair. Hard negatives mean that the identities of HR-LR resolutions both violate two triplet constraints in Eq 1 and 2, respectively, such that

$$\begin{aligned} &\|f_h(\mathbf{x}_{hi}^{sa}) - f_l(\mathbf{x}_{li}^{tn})\|_2^2 - \|f_h(\mathbf{x}_{hi}^{sa}) - f_l(\mathbf{x}_{li}^{sp})\|_2^2 < \beta, \\ &\|f_l(\mathbf{x}_{li}^{sa}) - f_h(\mathbf{x}_{hi}^{tn})\|_2^2 - \|f_l(\mathbf{x}_{li}^{sa}) - f_h(\mathbf{x}_{hi}^{sp})\|_2^2 < \gamma. \end{aligned}$$

Unlike the triple loss for FR that needs a large batch size to ensure a minimal number of exemplars of any one identity occurred in each mini-batch, the TTL can keep stability and faster convergence with smaller mini-batches since there exist sufficient anchor-positive pairs in all HR-LR mini-batch pairs.

2.6. Resolution-specific Discriminative Learning

In addition to adapting the model to be resolution-invariant, we also adopt the joint supervision of softmax loss L_s^* and center loss L_c^* [2] to ensure the resolution-specific discriminability of the TCN. The joint loss $L_d^* = L_s^* + \lambda L_c^*$ in a general form for both domains, with $* \in \{h, l\}$ denoting the HR or LR domain, is defined as:

$$L_d^* = -\sum_{i=1}^{N} \log \frac{e^{\mathbf{W}_{y_i}^T \mathbf{v}_i + b_{y_i}}}{\sum_{j=1}^{M} e^{\mathbf{W}_j^T \mathbf{v}_i + b_j}} + \lambda \sum_{i=1}^{N} \left\| \mathbf{v}_i - \mathbf{c}_{y_i}^v \right\|_2^2$$

where N is the number of training samples and M indicates the number of the identities in the training data. $\mathbf{v}_i$ refers to the feature representation extracted by HR-net or LR-net from i-th image $\mathbf{x}_i$. $\mathbf{W}$ and $\mathbf{b}$ are the weights of the softmax layer. $\mathbf{W}_j$ denotes the jth column vector of the $\mathbf{W}$. y_i is the class label for the ith sample and $\mathbf{c}_{y_i}^v$ denotes the y_ith class center of deep features $\mathbf{v}$. It is helpful to note that we only use L_d^h for pre-training the HR-net.

2.7. Joint Training

Combining the losses we introduced before, we constitute the overall loss for the TCN model as:

$$L_{total} = L_s^l + \lambda L_c^l + \rho L_{ttl}.$$

where λ and ρ are two scaling factors used for balancing three loss functions. In this way, we guarantee both the discriminability and resolution invariance of the LR features to match with HR gallery images during the joint training.

3. EXPERIMENT

3.1. Experimental Setup

Dataset The CASIA-WebFace [21] dataset is used as the training set to train the proposed TCN model. It consists of 494,414 images of 10575 subjects, which contain at least 14 images per subject. The face images are cropped and aligned to 112x96 pixels by affine transformation with facial landmarks detected by MTCNN [22]. Extensive experiments are conducted on popular LFW [23] and SCface [24] benchmarks to evaluate the proposed TCN model.

Network Architecture We evaluate the proposed TCN on two kinds of CNN networks, VGGFace [20], and ResNet [2]. It would be more convincing that the efficacy of TTL does not depend on any particular architecture.

Implemenation Details The margins β and γ are both set to be 0.1. The scaling factors λ and ρ are 0.008 and 0.1, respectively. The batch size for both domains is 150. The embedding size d of each image is 1024. We use Adam [25] for

Table 1. Face recognition accuracy (%) of different methods using different probe sizes on LFW dataset.

Probe size	8x8	12x12	16x16	20x20	112x96
NA-VGGFace [20]	75.0	82.6	89.3	93.4	97.7
VDSR-VGGFace [10]	73.6	83.5	88.6	94.0	
DRRN-VGGFace [9]	74.2	84.2	88.5	93.8	
FT-VGGFace	82.3	88.6	92.7	94.8	
DCR-VGGFace [17]	83.7	88.9	93.1	95.2	
TCN-VGGFace	**85.8**	**91.2**	**95.4**	**96.5**	
NA-ResNet [2]	72.7	84.1	92.3	95.4	98.8
VDSR-ResNet [10]	70.4	85.5	91.9	96.0	
DRRN-ResNet [9]	70.6	86.2	91.8	95.8	
FT-ResNet	88.9	93.8	95.9	96.8	
Trunk network [17]	88.2	91.6	95.5	96.8	
DCR-ResNet [17]	89.3	93.2	96.6	97.3	
TCN-ResNet	**90.5**	**94.7**	**97.2**	**97.8**	

optimizer with the initial learning rate 0.01. The maximum training epoch is 60. Each pixel of images is normalized to [-1.0, 1.0]. The hyperparameters are tuned on 10% randomly sampled held-out training data of the CASIA-WebFace.

3.2. Performance on LFW

The LFW dataset contains 13,233 images of 5749 subject, which has been extensively studied for unconstrained FR in recent years. Following the evaluation protocol in [23], we compute the mean verification accuracy by the ten-fold cross-validation scheme. Face images are cropped and aligned using same methods as on CASIA-WebFace images. For two images in the cross-resolution face verification paradigm, we take one as HR (112x96) gallery image and down-sample the other one to 8x8, 12x12, 16x16, or 20x20, and then up-sample it as the LR probe image (112x96). Same pipeline is used on CASIA-WebFace during the training. Cosine distance is used to calculate the similarity between two features. We compare with No adaptation (NA), Fine-tuning (FT), Trunk network [17] and DCR [17] on two different base models, VGGFace [20] and ResNet [2]. NA uses a CNN trained on HR images without any adaptation. FT advances the NA with further fine-tuning the network on LR images. Trunk network is trained with images of different resolutions. DCR adopts coupled-mappings to minimize the distance between the HR-LR pairs. We also compare with SR-based methods, VDSR [10] and DRRN [9], which are used to recover the probe images for testing. The experimental results are shown in Table 1. In the last column, the accuracies for HR probe images of the same resolution as gallery images are also presented. From the results, our approach shows significant and consistent improvements over the baselines on both two architectures. Our model outperforms SR-based methods, VDSR [10] and DRRN [9] by 8.32% and 8.12% on average respectively since they are not optimized for recognition purposes. Besides, our method outperforms the best baseline DCR by 1.48% on average, which demonstrates the effectiveness of incorporating the inter-class information.

Table 2. Face rates (%) of different methods at difference distances on SCFace dataset.

Distance	d1	d2	d3
MDS [15, 26]	60.3	66.0	69.5
DMDS [16]	61.5	67.2	62.9
LDMDS [16]	62.7	70.7	65.5
RICNN [27]	23.0	66.0	74.0
VGGFace [20]	41.3	75.5	88.8
FT-VGGFace	46.3	78.5	91.5
DCR-VGGFace [17]	62.3	91.0	94.8
TCN-VGGFace	**64.8**	**92.8**	**96.5**
ResNet [2]	36.3	81.8	94.3
FT-ResNet	54.8	86.3	95.8
Trunk network [17]	52.0	89.5	96.3
DCR-ResNet [17]	73.3	93.5	98.0
TCN-ResNet	**74.6**	**94.9**	**98.6**

3.3. Performance on SCFace

The SCFace is a real-world dataset, which contains images of 130 subjects captured by surveillance cameras under unconstrained indoor environment. For each subject, there are 15 images captured by surveillance cameras at three distances (five images at each distance), 4.20 m (d1), 2.60 m (d2), and 1.00 m (d3), and one mugshot image taken by a digital camera. Following the setting in [16], frontal mugshot images are regared as gallery images and images captured by surveillance cameras at distance $d_i, i = 1, 2, 3$ are used as probe images. We take CASIA-WebFace images of size 112x96 as HR images and those of 112x96, 30x30, and 20x20 as LR images for training of the TCN at distance of d3, d2, and d1, respectively. For the SCFace dataset, 50 out of 130 subjects are randomly chosen for fine-tuning and rest of the subjects are for testing. As such, there is no identity overlap between the training and testing sets. The same face images from CASIA-WebFace and SCFace datasets are used for the fine-tuning of VGGFace and ResNet models. The nearest-neighbor classifier is used to classify all probe images. We compare with the state-of-the-art CRFR methods, MDS [15, 26], DMDS [16], LDMDS [16], RICNN [27], No adaptation (NA), Fine-tuning (FT), Trunk network [17] and DCR [17]. As we can see from Table 2, the proposed TCN model also significantly outperforms the baselines on both two architectures, especially at exceedingly low resolution. Especially, our method outperforms the best baseline DCR by 1.88% on average.

4. CONCLUSION

In this paper, we propose a novel end-to-end Transferable Coupled Network (TCN) for cross-resolution face recognition. The proposed TTL can well address the resolution mismatch problem based on the selected cross-resolution triplets. Besides, an online triplet selection method is introduced to make the model more efficient and stable. Extensive experiments on public LFW and SCFace datasets empirically demonstrate the effectiveness of the proposed TCN model.

5. REFERENCES

[1] Florian Schroff, Dmitry Kalenichenko, and James Philbin, "Facenet: A unified embedding for face recognition and clustering," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 815–823.

[2] Yandong Wen, Kaipeng Zhang, Zhifeng Li, and Yu Qiao, "A discriminative feature learning approach for deep face recognition," in *European Conference on Computer Vision*. Springer, 2016, pp. 499–515.

[3] Weiyang Liu, Yandong Wen, Zhiding Yu, Ming Li, Bhiksha Raj, and Le Song, "Sphereface: Deep hypersphere embedding for face recognition," in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, vol. 1, p. 1.

[4] Hao Wang, Yitong Wang, Zheng Zhou, Xing Ji, Zhifeng Li, Dihong Gong, Jingchao Zhou, and Wei Liu, "Cosface: Large margin cosine loss for deep face recognition," *arXiv preprint arXiv:1801.09414*, 2018.

[5] Jiankang Deng, Jia Guo, and Stefanos Zafeiriou, "Arcface: Additive angular margin loss for deep face recognition," *arXiv preprint arXiv:1801.07698*, 2018.

[6] Himanshu S Bhatt, Richa Singh, Mayank Vatsa, and Nalini K Ratha, "Improving cross-resolution face matching using ensemble-based co-transfer learning," *IEEE Transactions on Image Processing*, vol. 23, no. 12, pp. 5654–5669, 2014.

[7] Zhifei Wang, Zhenjiang Miao, QM Jonathan Wu, Yanli Wan, and Zhen Tang, "Low-resolution face recognition: a review," *The Visual Computer*, vol. 30, no. 4, pp. 359–386, 2014.

[8] Zhangyang Wang, Shiyu Chang, Yingzhen Yang, Ding Liu, and Thomas S Huang, "Studying very low resolution recognition using deep networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 4792–4800.

[9] Ying Tai, Jian Yang, and Xiaoming Liu, "Image super-resolution via deep recursive residual network," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, vol. 1, p. 5.

[10] Jiwon Kim, Jung Kwon Lee, and Kyoung Mu Lee, "Accurate image super-resolution using very deep convolutional networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 1646–1654.

[11] Wilman WW Zou and Pong C Yuen, "Very low resolution face recognition problem," *IEEE Transactions on Image Processing*, vol. 21, no. 1, pp. 327–340, 2012.

[12] Maneet Singh, Shruti Nagpal, Mayank Vatsa, Richa Singh, and Angshul Majumdar, "Identity aware synthesis for cross resolution face recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2018, pp. 479–488.

[13] Tomer Peleg and Michael Elad, "A statistical prediction model based on sparse representations for single image super-resolution," *IEEE transactions on image processing*, vol. 23, no. 6, pp. 2569–2582, 2014.

[14] Muwei Jian and Kin-Man Lam, "Simultaneous hallucination and recognition of low-resolution faces based on singular value decomposition," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 25, no. 11, pp. 1761–1772, 2015.

[15] Sivaram Prasad Mudunuri and Soma Biswas, "Low resolution face recognition across variations in pose and illumination," *IEEE transactions on pattern analysis and machine intelligence*, vol. 38, no. 5, pp. 1034–1040, 2016.

[16] Fuwei Yang, Wenming Yang, Riqiang Gao, and Qingmin Liao, "Discriminative multidimensional scaling for low-resolution face recognition," *IEEE Signal Processing Letters*, vol. 25, no. 3, pp. 388–392, 2018.

[17] Ze Lu, Xudong Jiang, and Alex ChiChung Kot, "Deep coupled resnet for low-resolution face recognition," *IEEE Signal Processing Letters*, 2018.

[18] Zhao Zhang, Yun-Hao Yuan, Xiao-Bo Shen, and Yun Li, "Low resolution face recognition and reconstruction via deep canonical correlation analysis," in *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2018, pp. 2951–2955.

[19] Bin Liu, Yue Cao, Mingsheng Long, Jianmin Wang, and Jingdong Wang, "Deep triplet quantization," *MM, ACM*, 2018.

[20] Omkar M Parkhi, Andrea Vedaldi, Andrew Zisserman, et al., "Deep face recognition.," in *BMVC*, 2015, vol. 1, p. 6.

[21] Dong Yi, Zhen Lei, Shengcai Liao, and Stan Z Li, "Learning face representation from scratch," *arXiv preprint arXiv:1411.7923*, 2014.

[22] Kaipeng Zhang, Zhanpeng Zhang, Zhifeng Li, and Yu Qiao, "Joint face detection and alignment using multitask cascaded convolutional networks," *IEEE Signal Processing Letters*, vol. 23, no. 10, pp. 1499–1503, 2016.

[23] Gary B Huang, Marwan Mattar, Tamara Berg, and Eric Learned-Miller, "Labeled faces in the wild: A database forstudying face recognition in unconstrained environments," in *Workshop on faces in'Real-Life'Images: detection, alignment, and recognition*, 2008.

[24] Mislav Grgic, Kresimir Delac, and Sonja Grgic, "Scface–surveillance cameras face database," *Multimedia tools and applications*, vol. 51, no. 3, pp. 863–879, 2011.

[25] Diederik P Kingma and Jimmy Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[26] Soma Biswas, Gaurav Aggarwal, Patrick J Flynn, and Kevin W Bowyer, "Pose-robust recognition of low-resolution face images," *IEEE transactions on pattern analysis and machine intelligence*, vol. 35, no. 12, pp. 3037–3049, 2013.

[27] Dan Zeng, Hu Chen, and Qijun Zhao, "Towards resolution invariant face recognition in uncontrolled scenarios," in *Biometrics (ICB), 2016 International Conference on*. IEEE, 2016, pp. 1–8.