Identity-Aware Face Super-Resolution for Low-Resolution Face Recognition

Jin Chen, Jun Chen, *Member, IEEE*, Zheng Wang, *Member, IEEE*, Chao Liang, and Chia-Wen Lin, *Fellow, IEEE*

***Abstract*—Although deep learning-based face recognition techniques have achieved amazing performance in recent years, low-resolution (LR) face recognition remains challenging. In this letter, we address this problem by proposing an identity-aware face super-resolution network to recover identity information of LR faces. To learn identity-aware features effectively, the identity features are explicitly disentangled to two orthogonal components: the magnitude and angle of features that project identity features to a hypersphere space. We show that the magnitude of features is related to the quality of a face. The proposed approach shows its superiority on recovering identity-related textures which are beneficial to recover identity information for recognition. Extensive experiments demonstrate the effectiveness of the proposed algorithm in LR face recognition.**

***Index Terms*—Low-resolution face recognition, face super-resolution, identity-aware learning, magnitude loss.**

I. Introduction

DEEP learning-based face recognition algorithms [1]–[7] have achieved amazing performances on public face datasets such as LFW [8]. But directly applying these models to low-resolution face recognition (LRFR) tasks usually degrades recognition performance severely, since they are trained on high-resolution (HR) faces which have clear structural information and abundant identity details, rather than on low-resolution (LR) faces that lack such information. The empirical study in [9] reveals that the performance drops significantly for LRFR when the resolution of a face is less than 32×32 pixels. Particularly, in many practical surveillance scenarios, people may be far away from cameras, thereby making the captured faces rather small. Hence, matching a LR face image against a HR gallery of the enrolled face like LRFR is a desirable and challenging task.

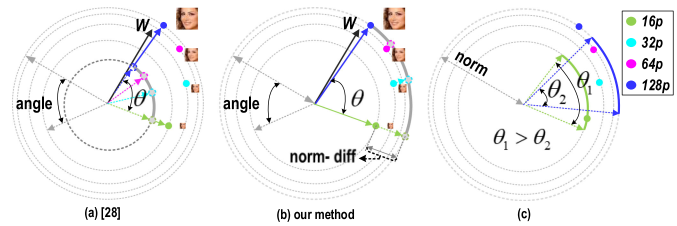


Fig. 1. Illustrations of different embedding strategies for identity embedding learning: (a) the angle-based embedding strategy [28]; (b) the proposed method; (c) the features with a smaller magnitudes are more sensitive to feature perturbation in the cosine metric space.

Algorithms for LRFR task can be mainly categorized into two approaches: methods that learn a unified feature space [10]–[15] and super-resolution (SR) based methods [16]–[22]. In this letter, we focus on the latter approach which reconstructs a HR face from a LR input for better face recognition. Numerous SR-based approaches have achieved great success in recovering HR details. For example, in [18] a discriminative generative network was proposed to generate authentic face images for ultra-resolving face images. Besides, SRGAN proposed in [20] can generate photo-realistic HR images. Nevertheless, these works mainly focus on generating HR faces in human visual perception rather than recovering identity-related details explicitly. Recently, some works aim at identity-oriented face SR [23]–[28]. [24] proposed an end-to-end deep convolution network with a hallucination sub-network cascaded by a recognition sub-network. Most recently, [28] proposed SICNN to enhance the identity information in reconstructed HR faces.

As shown in Fig. 1(a), SICNN projects the identity features of various resolutions into a unit hypersphere identity metric space (the hollow dots) and penalizes the normalized Euclidean distance between the reconstructed HR faces and the original HR faces. As such, it only focuses on the angle discrepancy of the identity-aware features. Although this angle-based identity embedding strategy can effectively improve LRFR performance, it tends to be sensitive to feature perturbations. For example, Fig. 1(c) shows the same stride (the green arc and blue arc) on different-magnitude equipotential lines, indicating that a smaller radius results in a larger angle change ($\theta_1 > \theta_2$). In other words, those features with smaller magnitudes are more sensitive to feature perturbations in the cosine metric space. To enhance the robustness of identity feature learning, as shown in Fig. 1(b), we propose to disentangle the identity features to two orthogonal

Manuscript received August 26, 2019; revised January 5, 2020; accepted January 28, 2020. Date of publication April 20, 2020; date of current version May 15, 2020. This work was supported in part by the National Key R&D Program of China under Grant 2017YFC0803700, in part by the National Nature Science Foundation of China under Grant 61876135, in part by the Nature Science Foundation of Hubei Province under Grant 2019CFB472, and in part by the Hubei Province Technological Innovation Major Project under Grant 2018AAA062 and Grant 2018CFA024. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Vasileios Mezaris. *(Corresponding author: Jun Chen.)*

Jin Chen, Jun Chen, and Chao Liang are with the National Engineering Research Center for Multimedia Software, School of Computer, Wuhan University, Wuhan 430072, China, and also with the Hubei Key Laboratory of Multimedia and Network Communication Engineering, Wuhan University, Wuhan 430072, China (e-mail: chenjin_2046@whu.edu.cn; chenj.whu@gmail.com; cliang@whu.edu.cn).

Zheng Wang is with the Digital Content and Media Sciences Research Division, National Institute of Informatics, Tokyo 101-8430, Japan (e-mail: wangz@nii.ac.jp).

Chia-Wen Lin is with the Department of Electrical Engineering, National Tsinghua University, Hsinchu 30013, Taiwan (e-mail: cwlin@ee.nthu.edu.tw).

Digital Object Identifier 10.1109/LSP.2020.2986942

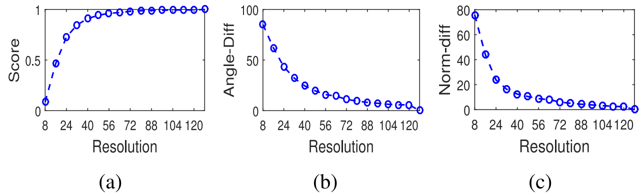


Fig. 2. Impacts of face resolution on feature representation: (a) the similarity score between HR-LR image pair decreases as the resolution decreases, especially when the resolution is 32 pixels; (b) the angle gap between HR-LR pair decreases with the face resolution; (c) the feature magnitude discrepancy also decreases with the face resolution.

components explicitly: the angle and magnitude of features. The aim of the identity embedding learning is not only to minimize the angle discrepancy of features between the reconstructed HR face and the original face in the angle space but also to promote the magnitude of the features of the reconstructed HR faces to make them more robust to feature perturbations. The main contributions of this letter are as follows: (i) We propose an identity-aware face SR network that not only reconstructs a high-fidelity HR appearance from a LR face but also recovers the face's identity information. (ii) We propose a novel feature decoupling strategy to make full use of both the angle and magnitude discrepancies of HR-LR face pairs for effectively learning the identity-aware features. (iii) We propose a novel magnitude loss to alleviate the magnitude discrepancy of features between the reconstructed and original HR faces.

II. Proposed Method

A. Intuition and Motivation

To understand the impact of face resolution on feature representation learning, inspired by [29], we evaluate the effect of down-scaling faces on the deep representations extracted by using the LightCNN_v9 model [6]. We calculate the mean cosine similarity between the feature pairs of each of 2,000 HR faces randomly sampled from VGGFace2 [30] and its corresponding downscaled vision with the description setting in [29]. As shown in Fig. 2(a), reducing the face resolution weakens the ability of deep representations in identity recognition. Notably, when the face resolution is less than 32×32 pixels, the performance drops rapidly. This discourages directly applying state-of-the-art face recognition models that are trained on high-resolution face dataset to LRFR, especially when the face resolution is low scale.

Motivated by [31], we propose to disentangle identity features into the angle and magnitude components of features for better understanding identity representations as explained in Fig. 2(b)-(c). Fig. 2(b) shows that the lower the face resolution is, the larger the average angle discrepancy between the features of HR-LR face pairs is. Meanwhile, as shown in Fig. 2(c), the average magnitude discrepancy of features increases with the face resolution decreases. In other words, besides the angle discrepancy, the magnitude discrepancy of identity features can be used to predict the quality of a reconstructed HR face, depending on the face resolution. As a result, LR faces not only enlarge the angle of features but also widen the magnitude gap of features. Both the effects result in the degeneration of deep representations ultimately.

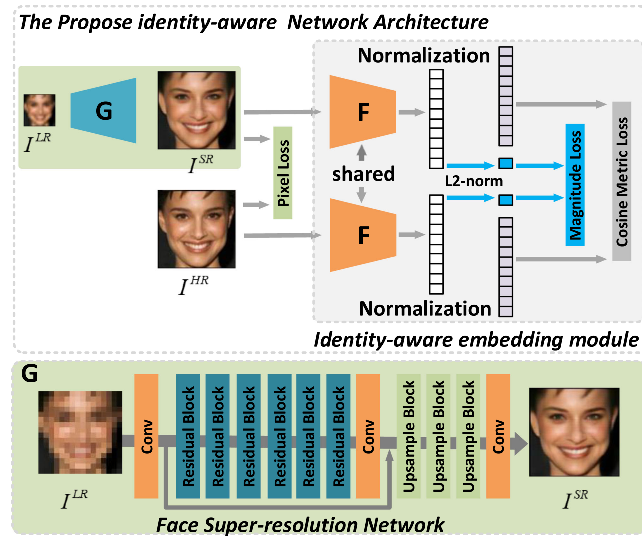


Fig. 3. The upper part shows the proposed identity-aware face SR network consisting of a face super-resolution network (denoted as G) and an identity-aware embedding module that contains a feature extractor (denoted as F). G reconstructs a HR face from a LR input face, and F extracts the identity features of the reconstructed HR face, and then the features are disentangled to the angle and magnitude of features for identity-embedding learning efficient. The lower pipeline shows the architecture of the proposed face super-resolution network.

Based on the key observation mentioned above, we propose a feature decoupling learning strategy to concern the angle discrepancy and the magnitude discrepancy simultaneously where the identity features are explicitly disentangled into two orthogonal components: the identity-related features (angle) and quality-related feature (magnitude) to recover identity-aware details.

B. Face Super-Resolution Network Architecture

The key of LRFR is to extract the features that are robust to resolution changes. As shown in Fig. 3, we propose an identity-aware face SR network pipeline aiming to reconstruct a high-fidelity and identity-aware HR face. The overall network contains two sub-modules: the face SR (FSR) module and the identity-aware embedding (IAE) module. The FSR module aims to hallucinate a HR face I^{SR} from a LR input I^{LR}, whereas the IAE module projects both the reconstructed HR face I^{SR} and the original HR face I^{HR} to the identity feature space. In the feature space, the identity features are disentangled into two components: the angle and magnitude of features. We then design loss functions to guide the network training to jointly optimize the visual fidelity and identity information of reconstructed HR faces. The goal of the network is to reconstruct a high-fidelity HR face in the pixel space, as well as minimize the angle and magnitude discrepancies of the features between a reconstructed HR face and its original HR face. We summarize the two modules as follows: $I^{SR} = G(I^{LR})$, $F^{in} = F(I^{in})$. Where G and F denote the FSR and IAE modules, respectively, F^{in} denotes the extracted feature representations of $I^{in} \in \{I^{SR}, I^{HR}\}$.

The FSR module is mainly constructed of three types of blocks: the convolution block, residual block and upsampling block which are described in [20]. The network consists of six residual blocks followed by three upsampling blocks. Each residual block is constructed by cascading two

Conv+BN operators with a parametric ReLU (PReLU) in-between the two Conv+BN. Each upsampling block consists of Conv+PixelShuffle+PReLU in sequence and the upscaling factor is 2. The number of channels of the input and output layers is 3 (RGB) and the kernel size is 9×9. For the remaining convolutional layers, the feature maps have 64 channels with a kernel size of 3×3. The IAE module extracts the identity features of input faces using the pre-trained LightCNN model [6], and then the identity features are disentangled for better identity embedding learning.

C. Identity-Aware Loss

To utilize the identity information effectively for guiding the training process, the identity features are explicitly disentangled into two orthogonal components: the angle and magnitude of features. Then we use the cosine metric function as the angle-related loss to guide identity feature learning directly.

Cosine Metric Loss: For a LR face input I_i^{LR}, we calculate the cosine distance $\cos\theta$ between the features of the reconstructed HR face I_i^{SR} and it's corresponding ground-truth I_i^{HR} in the angle space as follows:

$$L_{i,c} = 1 - \cos\theta, \tag{1}$$

where $\cos\theta = \frac{(F_i^{SR})^T (F_i^{HR})}{\|F_i^{SR}\|_2 \|F_i^{HR}\|_2}$, F_i^{SR} and F_i^{HR} are the identity feature vectors extracted by face recognition model *F* for faces I_i^{SR} and I_i^{HR}, respectively, θ is the angle between the feature vectors of I_i^{SR} and I_i^{HR}, and $\|F_i^{SR}\|_2$ and $\|F_i^{HR}\|_2$ denote the L2-norm of feature representations. As represented in (1), because the cosine distance between feature F_i^{SR} and F_i^{HR} ranges in $[-1, 1]$, we simply take its positive value.

Magnitude Loss: As shown in Fig. 2(c), a LR face not only enlarges the angle discrepancy between the features of the HR-LR pair but also increases the magnitude discrepancy of features. However, the cosine metric loss only concerns the angle discrepancy. As explained in Fig. 1(c), features with smaller magnitude values are more sensitive to feature perturbations in the hypersphere identity metric space. To take full advantage of the magnitude of HR face supervision information to learn robust identity representations, we propose a magnitude loss which pays more attention to the quality of reconstructed face to alleviate the magnitude discrepancy between features F_i^{SR} and F_i^{HR} as defined below:

$$L_{i,a} = \left\|\mathrm{norm}(F_i^{SR}) - \mathrm{norm}(F_i^{HR})\right\|_2 \tag{2}$$

where $\mathrm{norm}(F_i) = \|F_i\|_2$, $L_{i,a}$ measures the gap between F_i^{SR} and F_i^{HR} in the magnitude domain.

D. Pixel-Wise Loss

For face SR network training, we use the Euclidean distance to measure the content loss of the super-resolution network to ensure the reconstructed visual fidelity.

$$L_{i,p} = \left\|I_i^{SR} - I_i^{HR}\right\|_2^2. \tag{3}$$

Overall loss function: The overall loss function for training the proposed identity-aware face SR network consists of the pixel-wise loss, cosine metric loss, and magnitude loss:

$$L_{\text{total}} = \frac{1}{m}\sum_{i=1}^{m}(\alpha L_{i,p} + \beta_1 L_{i,c} + \beta_2 L_{i,a}), \tag{4}$$

where *m* is the batch size of training, α, β_1 and β_2 are the weights to balance the pixel-wise loss, cosine metric loss, and magnitude loss, respectively. In the experiments, we set parameters $\alpha = 1$, $\beta_1 = 10$, $\beta_2 = 1e^{-3}$.

III. Experiments

We evaluate the proposed method on LFW [8] and CelebA [32], which are widely-used benchmarks for face recognition and face SR in unconstrained environments, respectively. The entire training procedure contains two stages. At first, we train the FSR module with the pixel-wise loss solely since the qualities of reconstructed HR faces are too poor to learn identity-aware features effectively at the early epoch of training. We randomly sample about 190K face images from VGGFace2 [30] as the training set, where 22 faces per identity are selected. Secondly, we fine-tune the FSR module by joint supervision of the pixel-loss and the identity-aware loss. We use a pre-trained LightCNN model [6] which train on the entire VGGFace2 as the identity feature extractor and fix the parameters to focus on the identity-aware feature learning of the FSR module. All face images in the training and testing stages are normalized to 128×128 pixels where faces are detected and aligned by MTCNN [33]. We train the FSR module for 80 epochs with a learning rate of $2e^{-3}$, and jointly train the FSR module and the IAE module for another 40 epochs with a learning rate of $1e^{-4}$. The batch size is 128. We only perform data augmentation by random horizontal flipping with a probability of 0.5. At the testing stage, we use the cosine similarity score as the metric. The proposed method is compared with several representative methods based on different supervisions, including SR-net, SR-GAN [20], SR+Ou [19], SR+Angle [28] for the LRFR task. In the compared methods, SR-net denotes to train the FSR network guided by the pixel-wise loss only. In contrast, SRGAN emphasizes the use of adversarial loss [34]. In SR+Ou, the identity-aware loss is calculated by the Euclidean distances between feature pairs directly, whereas SR+Angle decouples the feature into the angle part and the magnitude part and only utilizes the angle discrepancy as identity signal.

A. Evaluation on Low-Resolution Face Recognition

Low-Resolution Face Verification (LRFV): We evaluate the performance of LRFV on LFW that aims to identify whether a LR-HR face pair belongs to the same person or not. We construct the LR-HR pair set by randomly selecting a face of pair of original 6,000 face pairs and down-scaling them from 128×128 to 16×16 via interpolation. We use two pre-trained face recognition models LightCNN-v9 and LightCNN-v29 [6] without fine-tuning on training set as an identity feature extractor. We evaluate all methods use the code provided in the link.[1] As shown in the third column of Table I, all face SR-based methods outperform bicubic interpolation, showing that SR can effectively improve face verification performance. The performance of SRGAN is the poorest among the face SR schemes as it does not explicitly learn identity-aware features. Although SR+Ou learns the identity-aware features by embedding identity features in the Euclidean space, which, however, often convolves the intra-class variations with the inter-class variations, making it hard to explicitly adjust individual variations for embedding learning. SR+Angle mainly tackles with the angle-related variations, but cannot fully address the problem. Compared with

[1] https://github.com/AlfredXiangWu/face_verification_experiment

TABLE I
PERFORMANCE EVALUATION ON LFW AND CELEBA. FOR VERIFICATION, ACC. AND VR@FAR ARE REPORTED. FOR IDENTIFICATION, RANK-1 ACCURACY IS PRESENTED

		LFW			CelebA
Model	Method	ACC.	VR@ FAR=1%	VR@ FAR=0.1%	Rank-1
LightCNN-v9	Bicubic	87.56%	32.10%	15.83%	1.64%
	SR-net	97.20%	75.60%	59.67%	23.69%
	SRGAN	97.03%	74.30%	52.40%	23.15%
	SR+Ou	98.16%	81.20%	69.43%	30.22%
	SR+Angle	98.35%	82.90%	69.23%	30.15%
	Proposed	**98.46**%	**83.73**%	**70.67**%	**30.39**%
LightCNN-v29	Bicubic	90.89%	43.67%	22.67%	3.24%
	SR-net	98.24%	80.63%	67.00%	31.77%
	SRGAN	98.01%	79.70%	62.37%	30.73%
	SR+Ou	98.87%	86.83%	75.60%	39.23%
	SR+Angle	98.96%	**88.37**%	76.07%	40.54%
	Proposed	**98.98**%	87.70%	**76.67**%	**41.36**%

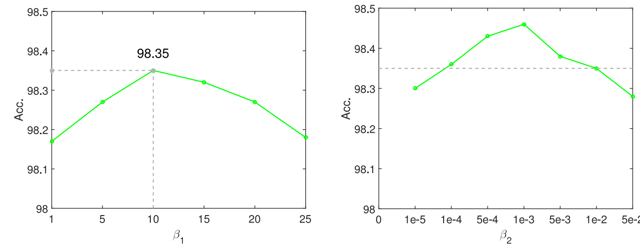
SR+Angle, the proposed method takes into account both the angle-related and magnitude-related discrepancies of features simultaneously. With the proposed joint guidance with pixel-wise loss and disentangled magnitude and angle identity loss functions on network training, our method achieves the best verification accuracy of 98.46% and 98.98% on two identity models respectively. We also evaluate the performance based on a more practical metric, i.e., the face verification rate at a low false acceptance rate. The proposed method achieves 83.73% VR@FAR = 1% and 70.67% VR@FAR = 0.1% on LightCNN-v9, outperforming the other methods.

Low-Resolution Face Identification (LRFI): We evaluate the LRFI performance with Rank-1 on CelebA [32] that contains 202,599 faces of 10,177 identities with the large pose, occlusion, and expression variations. For those identities with more than 25 samples, we select a high-quality frontal face for each of the identities to constitute the gallery set, and their rest faces are included in the probe set after 8× down-sampling. In total, the gallery set and the probe set contain 3,661, 102,678 faces, respectively. As shown in the last column of Table I, the proposed method achieves 30.39% and 41.36% at Rank-1. Again, the proposed method outperforms other face FR schemes.

B. Ablation Study

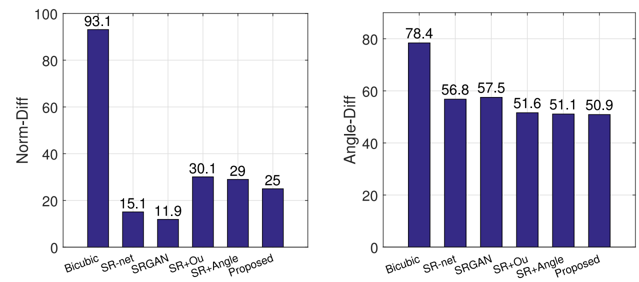
Balancing the magnitude and angle signals: Because the identity-aware loss contains two parts: cosine metric loss and magnitude loss, we investigate the interactions of the magnitude and angle signals on the identity embedding learning. Inspired by [28], we first set the parameter of magnitude loss $\beta_2 = 0$ and optimize the parameters α and β_1. We set $\alpha = 1$ and search the optimal parameter $\beta_1 \in (1, 25)$ with stride 5. As shown in Fig. 4(a), when β_1 near 10, we achieve the best verification accuracy with 98.35% on LFW. Then we fix $\alpha = 1$, $\beta_1 = 10$ as the baseline denoted by a gray dotted line in Fig. 4(b) and optimize the parameter β_2. We achieve a steady performance improvement when $\beta_2 \in$ (1e-4, 1e-2) and obtain 98.46% when β_2 is 1e-3.

The effect of embedding methods on the magnitude and angle gap: To understand different embedding methods working on the magnitude gap and angle gap between generated and original HR face pairs, we randomly select 2,000 face images from the



(a) (b)

Fig. 4. (a) Shows face verification accuracy (Acc.) by varying the weighting parameter β_1 with $\alpha = 1$, $\beta_2 = 0$, (b) shows face verification accuracy by varying the weighting parameter β_2 by fixing $\alpha = 1$, $\beta_1 = 10$.



(a) (b)

Fig. 5. (a) Shows the mean magnitude gap of features between generated HR face and original HR face, and (b) shows the mean angle gap of feature pairs.

VGGFace2 [30] dataset with no overlapping to the training set. Fig. 5 shows the mean angle and magnitude gap of generated and original HR face pairs. It shows that all SR-based methods reduce the angle and magnitude gap of features. Compared with SR-net, SRGAN pays more attention to reducing the magnitude gap of features than the angle gap of features. On the contrary, SR+Ou pays more attention to reducing the angle gap of features than the magnitude gap of features. Compared with SR+Ou, SR+Angle alleviates the angle gap with angle-related supervision, interestingly, it slightly reduces the magnitude gap without magnitude supervision explicitly. By decoupling features to the angle part and the magnitude part and supervising explicitly, the proposed method decreases the angle gap and the magnitude gap further.

IV. CONCLUSION

In this letter, we present an identity-aware face super-resolution network to recover the identity information of the generated facial image. Inspired by analyzing the impact of low-resolution on deep features, we decouple the identity features into the angle-related features and the magnitude-related features for supervision explicitly. Extensive experiments demonstrate the effectiveness of the proposed algorithm in the low-resolution face recognition.

ACKNOWLEDGMENT

The numerical calculations in this letter have been done on the supercomputing system in the Supercomputing Center of Wuhan University.

REFERENCES

[1] T. Chan, K. Jia, S. Gao, J. Lu, Z. Zeng, and Y. Ma, "PCANet: A simple deep learning baseline for image classification?," *IEEE Trans. Image Process.*, vol. 24, no. 12, pp. 5017–5032, Dec. 2015.

[2] J. Lu, J. Hu, and Y. Tan, "Discriminative deep metric learning for face and kinship verification," *IEEE Trans. Image Process.*, vol. 26, no. 9, pp. 4269–4282, Sep. 2017.

[3] G. Hu, X. Peng, Y. Yang, T. M. Hospedales, and J. Verbeek, "Frankenstein: Learning deep face representations using small data," *IEEE Trans. Image Process.*, vol. 27, no. 1, pp. 293–303, Jan. 2017.

[4] W. Liu, Y. Wen, Z. Yu, M. Li, B. Raj, and L. Song, "Sphereface: Deep hypersphere embedding for face recognition," in *Proc. IEEE Conf. Comput. Vision Pattern Recognit.*, 2017, pp. 212–220.

[5] Y. Wen, K. Zhang, Z. Li, and Y. Qiao, "A discriminative feature learning approach for deep face recognition," in *Proc. Eur. Conf. Comput. Vision*, 2016, pp. 499–515.

[6] X. Wu, R. He, Z. Sun, and T. Tan, "A light CNN for deep face representation with noisy labels," *IEEE Trans. Inf. Forensics Secur.*, vol. 13, no. 11, pp. 2884–2896, Nov. 2018.

[7] F. Schroff, D. Kalenichenko, and J. Philbin, "Facenet: A unified embedding for face recognition and clustering," in *Proc. IEEE Conf. Comput. Vision Pattern Recognit.*, 2015, pp. 815–823.

[8] T. Berg, G. B. Huang, M. Ramesh and E. Learned-Miller, "Labeled faces in the wild: A database for studying face recognition in unconstrained environments," Univ. Massachusetts, Amherst, MA, USA, Tech. Rep. 07–49, Oct. 2007.

[9] W. W. Zou and P. C. Yuen, "Very low resolution face recognition problem," *IEEE Trans. Image Process.*, vol. 21, no. 1, pp. 327–340, Jan. 2011.

[10] C. Ren, D. Dai, and H. Yan, "Coupled kernel embedding for low-resolution face image recognition," *IEEE Trans. Image Process.*, vol. 21, no. 8, pp. 3770–3783, Aug. 2012.

[11] Z. Lu, X. Jiang, and A. Kot, "Deep coupled resnet for low-resolution face recognition," *IEEE Signal Process. Lett.*, vol. 25, no. 4, pp. 526–530, Apr. 2018.

[12] F. Yang, W. Yang, R. Gao, and Q. Liao, "Discriminative multidimensional scaling for low-resolution face recognition," *IEEE Signal Process. Lett.*, vol. 25, no. 3, pp. 388–392, Mar. 2018.

[13] Y. Chu, T. Ahmad, G. Bebis, and L. Zhao, "Low-resolution face recognition with single sample per person," *Signal Process.*, vol. 141, pp. 144–157, Dec. 2017, doi: 10.1016/j.sigpro.2017.05.012.

[14] S. Ge, S. Zhao, C. Li, and J. Li, "Low-resolution face recognition in the wild via selective knowledge distillation," *IEEE Trans. Image Process.*, vol. 28, no. 4, pp. 2051–2062, Apr. 2019.

[15] P. Li, L. Prieto, D. Mery, and P. J. Flynn, "On low-resolution face recognition in the wild: Comparisons and new techniques," *IEEE Trans. Inf. Forensics Secur.*, vol. 14, no. 8, pp. 2000–2012, Aug. 2019.

[16] J. Jiang, Y. Yu, S. Tang, J. Ma, A. Aizawa, and K. Aizawa, "Context-patch face hallucination based on thresholding locality-constrained representation and reproducing learning," *IEEE Trans. Cybern.*, vol. 50, no. 1, pp. 324–337, Jan. 2020.

[17] E. Zhou, H. Fan, Z. Cao, Y. Jiang, and Q. Yin, "Learning face hallucination in the wild," in *Proc. AAAI Conf. Artif. Intell.*, 2015, pp. 3871–3877.

[18] X. Yu and F. Porikli, "Ultra-resolving face images by discriminative generative networks," in *Proc. Eur. Conf. Comput. Vision*, 2016, pp. 318–333.

[19] J. Justin, A. Alexandre, and F. F. Li, "Perceptual losses for real-time style transfer and super-resolution," in *Proc. Eur. Conf. Comput. Vision*, 2016, pp. 694–711.

[20] C. Ledig *et al.*, "Photo-realistic single image super-resolution using a generative adversarial network," in *Proc. IEEE Conf. Comput. Vision Pattern Recognit.*, 2017, pp. 4681–4690.

[21] J. Kim, K. L. Jung, and M. L. Kyoung, "Accurate image super-resolution using very deep convolutional networks," in *Proc. IEEE Conf. Comput. Vision Pattern. Recognit.*, 2016, pp. 1646–1654.

[22] Z. Wang, M. Ye, F. Yang, X. Bai, and S. Satoh, "Cascaded SR-GAN for scale-adaptive low resolution person re-identification," in *Proc. AAAI Conf. Artif. Intell.*, 2018, pp. 3891–3897.

[23] J. Wu, S. Ding, W. Xu, and H. Chao, "Deep joint face hallucination and recognition," 2016, *arXiv:1611.08091*.

[24] Z. Wang, S. Chang, Y. Yang, D. Liu, and T. S. Huang, "Studying very low resolution recognition using deep networks," in *Proc. IEEE Conf. Comput. Vision Pattern Recognit.*, 2016, pp. 4792–4800.

[25] W. Su, C. Hsu, C. Lin, and W. Lin, "Supervised-learning based face hallucination for enhancing face recognition," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, 2016, pp. 1751–1755.

[26] C. Hsu, C. W. Lin, W Su, and C. Gene, "Sigan: Siamese generative adversarial network for identity-preserving face hallucination," *IEEE Trans. Image Process.*, vol. 28, no. 12, pp. 6225–6236, Dec. 2019.

[27] Z. Cheng, X. Zhu, and S. Gong, "Low-resolution face recognition," in *Proc. Asian Conf. Comput. Vision*, 2018, pp. 605–621.

[28] K. Zhang *et al.*, "Super-identity convolutional neural network for face hallucination," in *Proc. Eur. Conf. Comput. Vision*, 2018, pp. 183–198.

[29] W. Tan, B. Yan, and B. Bare, "Feature super-resolution: Make machine see more clearly," in *Proc. IEEE Conf. Comput. Vision Pattern Recognit.*, 2018, pp. 3994–4002.

[30] Q. Cao, L. Shen, W. Xie, Q. M. Parkhi, and A. Zisserman, "VGGFace2: A dataset for recognising faces across pose and age," in *Proc. IEEE Int. Conf. Autom. Face Gesture Recognit.*, 2018, pp. 67–74.

[31] W. Liu *et al.*, "Decoupled networks," in *Proc. IEEE Conf. Comput. Vision Pattern. Recognit.*, 2018, pp. 2771–2779.

[32] Z. Liu, P. Luo, X. Wang, and X. Tang, "Deep learning face attributes in the wild," in *Proc. IEEE Conf. Comput. Vision Pattern Recognit.*, 2015, pp. 3730–3738.

[33] K. Zhang, Z. Zhang, Z. Li, and Y. Qiao, "Joint face detection and alignment using multitask cascaded convolutional networks," *IEEE Signal Process. Lett.*, vol. 23, no. 10, pp. 1499–1503, Oct. 2016.

[34] I. Goodfellow *et al.*, "Generative adversarial nets," in *Proc. Int. Conf. Neural Inf. Process. Syst.*, 2014, pp. 2672–2680.