# Unsupervised Face Domain Transfer for Low-Resolution Face Recognition

Sungeun Hong, *Member, IEEE*, and Jongbin Ryu, *Member, IEEE*

***Abstract*—Low-resolution face recognition suffers from domain shift due to the different resolution between a high-resolution gallery and a low-resolution probe set. Conventional methods use the pairwise correlation between high-resolution and low-resolution for the same subject, which requires label information for both gallery and probe sets. However, explicitly labeled low-resolution probe images are seldom available, and labeling them is labor-intensive. In this paper, we propose a novel unsupervised face domain transfer for robust low-resolution face recognition. By leveraging the attention mechanism, the proposed generative face augmentation reduces the domain shift at image-level, while spatial resolution adaptation generates domain-invariant and discriminant feature distributions. On public datasets, we demonstrate the complementarity between generative face augmentation at image-level and spatial resolution adaptation at feature-level. The proposed method outperforms the state-of-the-art supervised methods even though we do not use any label information of low-resolution probe set.**

***Index Terms*—Low-resolution face recognition, image-to-image translation, domain adaptation, attention, face augmentation.**

## I. Introduction

Although recent face recognition methods have shown promising results [1]–[3], heterogeneous domains between the gallery and probe sets still remains a challenging problem [4], [5]. A popular example of the heterogeneity is the significant difference between a high-quality gallery (*e.g.*, identification photos) and low-quality probe images from surveillance cameras [6]. Photos used in ID cards or passports are captured in a stable environment to register as gallery images. On the other hand, probe images are usually captured under unstable real environments including noise, blur, arbitrary pose as mentioned in [7], [8]. This causes severe domain shift between source (i.e., gallery) and target (i.e., probe) images, which degrades the recognition performance.

Considerable efforts [9]–[13] have been devoted to coping with the heterogeneous condition between the gallery and probe sets over the past decades. Among these studies, there has been an increasing interest in low-resolution face recognition (LRFR) [14], which takes probe images under low-quality conditions [15]–[17]. The existing methods can be divided into two categories. Hallucination approaches obtain high-resolution (HR) images from low-resolution (LR) probe images before performing face recognition [18]–[20]. On the other hand, models in the embedding category generate common feature space for HR and LR conditions [21]–[23]. Crucially, both approaches use identity labels for both HR and LR images. In other words, most existing LRFR methods learn the relations between HR gallery and LR probe sets through the supervision of identity labels.

However, we argue that unsupervised approaches that do not require the labeling of LR probe sets have significant advantages over supervised-based approaches. First of all, while it is not difficult to acquire LR images in advance, explicitly labeled images are seldom available and labeling them is quite labor-intensive and impractical. Second, the unsupervised setting has the advantage to improve the performance of the recognition model over time. Unsupervised face recognition systems do not require any label information of LR probe images, so they can easily evolve even the LR probe images are continuously given.

In this paper, we propose an unsupervised LRFR method to reduce domain shift at feature-level as well as image-level. Unlike previous approaches [18], [19], [21]–[23], our model uses only images without label information in an LR probe set. As can be seen in Fig. 1, the proposed method consists of generative face augmentation (GFA) and spatial resolution adaptation (SRA) networks to produce a discriminant

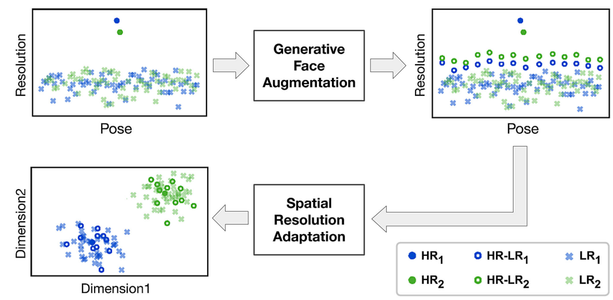


Fig. 1. Overview of the proposed method. The gallery set contains labeled high-resolution (HR) images, while the probe set includes low-resolution (LR) images without labels. Subscript ‘1’ or ‘2’ in the legend refers to the class label. Given HR images, generative face augmentation (GFA) synthesizes ‘HR-LR’ images so that they resemble LR images at the image-level. Spatial resolution adaptation (SRA) enables common feature spaces in both domains to be discriminative.

Manuscript received October 15, 2019; revised December 11, 2019; accepted December 19, 2019. Date of publication December 30, 2019; date of current version January 31, 2020. This work was supported by Basic Science Research Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Education (NRF-2017R1A6A3A11031193). The associate editor coordinating the review of this manuscript and approving it for publication was Dr. S. S. Channappayya. *(Corresponding author: Jongbin Ryu.)*

S. Hong is with T-Brain, AI Center, SK Telecom, 04539 Seoul, South Korea (e-mail: csehong@gmail.com).

J. Ryu is with the Department of Computer Science, Hanyang University, 04763 Seoul, South Korea (e-mail: jongbin.ryu@gmail.com).

Digital Object Identifier 10.1109/LSP.2019.2963001

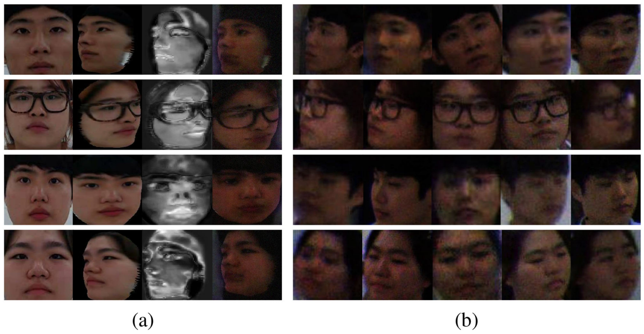

Fig. 2. Given labeled HR images and unlabeled LR images, our GFA can map two domains without any pair labels by leveraging attention mechanism. (a) (from left to right) HR gallery images, synthesized images with varying pose, learned attention map in which bright value indicates higher focus, and domain-transferred HR-LR images with various poses (b) LR probe images.

distribution that is aligned at both image and feature levels. The proposed GFA is designed to synthesize labeled HR images into LR-like images while SRA ensures that generated LR-like images are indistinguishable from real LR images at feature-level.

The proposed method is closely related to our previous work [24] in considering domain adaptation networks for heterogeneous face recognition. In our prior work [24], we generate synthetic images with a 3D face model, focusing only on pose variation. In contrast, this work introduces joint optimization of the learning-based generative network (i.e., GFA) with domain adaptation network (i.e., SRA) to generate domain-transferred images and also domain-invariant feature distribution. However, simply transferring images from HR domain to LR domain frequently results in mode collapse [25] and unexpected artifacts [26]. Therefore, our GFA includes the attention-guided domain transfer module to focus on critical parts for compensating the difference between the source and target images. Fig. 2 shows sample images in HR gallery images, domain-transferred images, and LR probe images.

We validate the proposed method in a challenging and realistic LRFR protocol on public datasets [24], [27], [28]. The gallery set includes HR images captured under stable environments, such as identification photos, while the probe set contains unlabeled LR images from surveillance cameras. We experimentally show that our proposed GFA and SRA perform complementarily and eventually show drastic performance improvement. Surprisingly, the accuracy of our unsupervised model is superior to the state-of-the-art supervised methods [10], [24], [29]–[31]. We also provide extensive ablation studies that can serve as quantitative baselines for domain adaptation-based approaches for LRFR.

## II. Related Works

**Low-Resolution Face Recognition (LRFR)**: Conventional LRFR methods can be grouped into two categories depending on the level at which resolution alignment is applied, i.e., image-level or feature-level. Early investigations in image-level methods include singular value decomposition [18] or sparse representation [19] for synthesizing person-specific low-resolution facial images. Hallucination-based reconstruction of HR images from input LR images was also studied. To transfer LR images into HR images, super-resolution techniques or facial attribute embedding methods have been widely used [11], [20], [32]. These methods transfer probe LR images into HR images, which cause over smoothness and loss of details. Conversely, we transfer HR images into LR images by using attention-guided domain transfer and also perform domain alignment at feature-level.

On the other hand, the goal of the feature-level category is to find a unified feature space where the proximity between LR and HR images is maintained for the same subject. Coupled mapping strategy is widely used for learning the projection between HR images and LR images [14], [21]. Dictionary-based approaches have also been suggested to match facial images captured at different resolutions [22], [23]. Crucially, most existing methods assume that there are both LR and HR versions available for each subject, which is impractical for real-world applications. In contrast, we propose an unsupervised method that does not require any labels in LR images.

**Domain Adaptation (DA):** Several attempts have been proposed to apply domain adaptation (DA) to face-related tasks. Xie *et al.* [33] use DA and several hand-crafted descriptors to tackle face recognition in which the gallery set consists of clear images while the probe set includes blurred images. Banerjee *et al.* [34] propose a DA-based method with a filter bank, *e.g.*, Eigenfaces, Fisherfaces, and Gaborfaces. Unlike the above approaches, which apply DA after extracting the handcrafted-features from images, we jointly perform feature learning and classification in an integrated deep architecture. Moreover, we solve the face recognition problem in which only one labeled image per person is given to the model, which is a challenging and realistic protocol.

## III. Proposed Method

### A. Generative Face Augmentation (GFA)

The goal of GFA is to learn a mapping function from source HR domain to target LR domain, i.e., $Gen_{HR \to LR}$. To bridge the gap between them, we use image-level domain transfers using 'pix2pix' framework [35], which consists of two sub-networks: 1) auto-encoder architecture called 'generator' that attempts to translate an image 2) 'discriminator' trying to classify whether the image is actually translated from 'generator.' Due to this generative adversarial relationship between them as in generative adversarial networks (GAN) [36], both generator and discriminator are getting better as learning progresses.

Since pair data of HR and LR images for the same subject is not available, we construct two generators and adopt a cycle consistency loss [37]. However, simply applying image-level domain transfer to the source and target domains can lead to unexpected artifacts and performance degradation. Motivated by the recent success of attention-based approaches in computer vision fields [38]–[40], we leverage attention mechanism during a domain transfer. To this end, we add two attention modules $Att_{HR->LR}$ and $Att_{LR->HR}$, which reduce the artifacts of synthesized images by focusing only on crucial parts in the

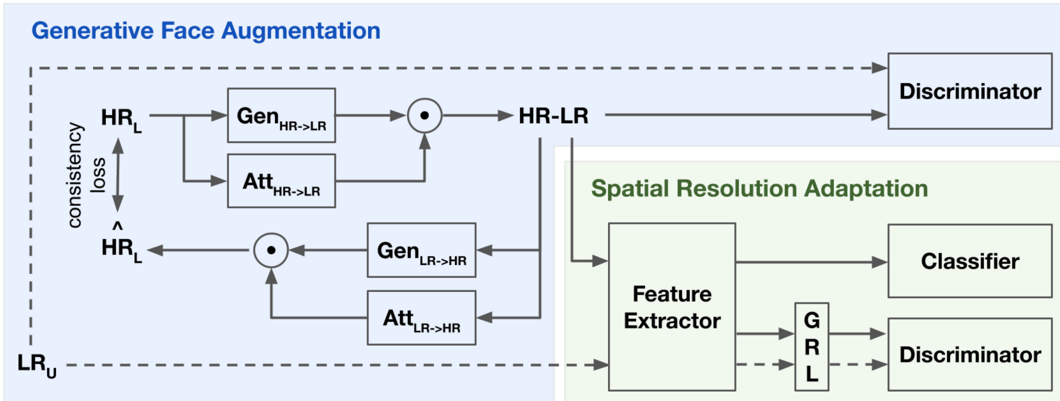


Fig. 3. Overall flow of the proposed method. Given labeled HR images, domain discriminator in generative face augmentation (GFA) encourages attention-guided generator to transfer them into LR-like images at image-level. Spatial resolution adaptation (SRA) ensures that LR-like images are indistinguishable from real LR images at feature-level. Note that the solid line is the path of the labeled HR images, while the dotted line is the path of the unlabeled LR images.

image, i.e., foreground. Given an HR image, we first feed it to the generator and attention module. We can generate the foreground area via a Hadamard product [41] between domain-transferred image and attention map. After that, we obtain the background area from the counterpart of the attention map, and add them to the image with the foreground transferred. As a result, we can obtain a domain transferred image as:

$$hr - lr = Gen_{HR \to LR}(hr) \odot m_{Att} + hr \odot (1 - m_{Att}), \quad (1)$$

where $hr$ is a image from source HR domain and $m_{Att}$ refers to attention map derived from attention module $Att_{HR->LR}$ containing the value $[0, 1]$ per pixel.

### B. Spatial Resolution Adaptation (SRA)

Although we generate image-level domain-transferred images, there still exists a difference between source HR domain and target LR domain at feature-level. Therefore, we introduce an SRA network that predicts class labels and domain labels (i.e., HR/LR) as can be seen in Fig. 3. SRA aims to learn the unified feature distribution that is discriminative in both HR and LR domains without LR domain labels. For the unified feature distribution, we update the parameters of $F$ and $C$, $\theta_F$ and $\theta_C$, to minimize the label prediction loss in HR domain. Here, $F$ indicates feature extractor and $C$ refers to the class-label classifier. Concurrently with the discriminant learning, we should align the distribution between HR domain and LR domain. In order to obtain domain-invariant features, SRA struggles to find a $\theta_F$ that maximizes the domain prediction loss, while simultaneously searching for parameters of domain discriminator $D$ ($\theta_D$) to minimize domain prediction loss as follows:

$$\begin{aligned} L &= \sum_{i \in HR} L_C^i + \sum_{i \in HR \cup LR} L_D^i \qquad \text{when update } \theta_D, \\ L &= \sum_{i \in HR} L_C^i - \lambda \sum_{i \in HR \cup LR} L_D^i \qquad \text{when update } \theta_F \text{ and } \theta_C, \end{aligned} \quad (2)$$

where, $L_C^i$ and $L_D^i$ refer to the loss of the $i$-th sample from class and domain predictions, respectively. The parameter $\lambda$ is the most critical factor in which a negative sign of $\lambda$ leads to an adversarial relationship between $F$ and $D$ in terms of loss. As a result, the parameters of $F$ converge at a compromise point that is discriminative and also satisfies domain invariance through the process of minimizing network loss.

## IV. Experimental Results

### A. Experimental Setup

To construct GFA, we build on the 'pix2pix' framework in which 'generator' consists of convolutions, residual blocks, and deconvolutions, and 'discriminator' is a two-class classification network with several convolutions. We apply several augmentation schemes (*e.g.*, flip, rotation, and crop) [42], [43] to prevent over-fitting and DA failure due to the lack of source domain images. Further, when pose variation is severe in the target domain, we optionally perform 3D image synthesis [24]. For SRA, we adopt pre-trained VGG-Face [1] as a feature extractor, attach shallow networks for classifier and discriminator, and finally fine-tune them. To prevent over-fitting on relatively small datasets, we freeze the weights from the Convl layer to FC6 layer of the pre-trained VGG-Face and fine-tune the FC7 layer. For more details of SRA, please refer to the project page[1] of our prior work [24]

We demonstrate the effectiveness of the proposed method on public datasets, EK-LFH [24], SCface [27], and YouTubeFaces (YTF) [28]. Details of each dataset are presented in Table I. Considering real-world scenarios, we follow the subject-disjoint protocol in LRFR [10], [29], [44], which prevents the use of a test image of subjects that appear in the training phase.

### B. Evaluation on EK-LFH

Following the subject-disjoint protocol [10], [29], [44] in EK-LFH [24], we use 20 randomly selected subjects for training

[1] https://github.com/csehong/SSPP-DAN

TABLE I
DATASET SPECIFICATION. TWO NUMBERS OF THE TARGET DOMAIN IN THE SUBJECT COLUMN INDICATE THE NUMBER OF SUBJECTS USED FOR TRAINING AND TEST, RESPECTIVELY

| Dataset | EK-LFH | | SCface | | YTF | |
|---|---|---|---|---|---|---|
| Domian | HR | LR | HR | LR | HR | LR |
| Subjects | 30 | 20/10 | 130 | 50/80 | 1595 | 798/797 |
| Samples | 30 | 16K | 130 | 2K | 3.5K | 306K/311K |

TABLE II
ABLATION ANALYSIS ON EK-LFH. SUBSCRIPTS 'L' AND 'U' INDICATE LABELED AND UNLABELED IMAGES AND 'POSE' REFERS TO SYNTHESIZED IMAGES WITH VARIOUS POSES

| Protocol | Training Set | Accuracy |
|---|---|---|
| Source Only | $\{HR_L\}$ | 20.97% |
| | $\{HR_L, HR_{pose}\}$ | 20.29% |
| | $\{HR_L, HR_{pose}, \text{HR-LR}\}$ | 26.81% |
| Proposed Adaptation | $\{HR_L, LR_U\}$ | 28.50% |
| | $\{HR_L, HR_{pose}, LR_U\}$ | 50.12% |
| | $\{HR_L, HR_{pose}, \text{HR-LR}, LR_U\}$ | 52.76% |
| Labeled Target | $\{HR_L, LR_L\}$ | 63.39% |

and the other 10 subjects for test. Pose variation is severe in the target domain on EK-LFH as can be seen in Table I, thus, we perform 3D image synthesis [24] during image-level domain transfer. Motivated by the experimental settings in conventional unsupervised DA [24], [45], we perform experiments on three protocols as shown in Table II. First, 'Source Only' protocol only uses labeled samples in the source HR domain, which revealed the theoretical lower bound on performance as 20.97%. We also report the results of adding our synthesized images (*e.g.*, $HR_{pose}$ or HR-LR) to the training set as variants of conventional 'Source Only.' Second, as an upper performance bound, the model from 'Labeled Target' protocol is trained by the labeled source HR images as well as labeled target LR images.

Finally, our main approaches, 'Proposed Adaptation,' use labeled source images with unlabeled target images. We observe that the proposed method with synthesized images improves accuracy compared to 'Source Only,' even though the labels of the target domain are not used. The fifth and sixth rows validate the importance of synthesized images with various poses when applying unsupervised DA. Also, we confirm that the proposed image-level domain transfer can improve performance even further. Overall results show that GFA (image-level) and SRA (feature-level) work complementarily in solving the LRFR task.

### C. Evaluation on SCface

Following the existing protocol in SCface [9], [10], [29], we split the training set and test set into 50 and 80 subjects separately. From Table III, we confirm a considerable performance improvement by using 'Proposed Adaptation' even if no label information of the target domain is used. It is apparent that the experimental results on SCface are consistent with that of the EK-LFH dataset. Finally, we compare our method with the state-of-the-art supervised methods that use labels in the target domain, as can be seen in Table IV. Despite not using any label information in the target LR domain, the proposed method outperforms the supervised LRFR methods. Finally, we can see that the proposed method including image-level domain transfer is superior to our previous work based on unsupervised DA [24].

TABLE III
ABLATION ANALYSIS ON SCface

| Protocol | Training Set | Accuracy |
|---|---|---|
| Source Only | $\{HR_L\}$ | 23.00% |
| | $\{HR_L, \text{HR-LR}\}$ | 26.08% |
| Proposed Adaptation | $\{HR_L, LR_U\}$ | 50.50% |
| | $\{HR_L, \text{HR-LR}, LR_U\}$ | 52.25% |

TABLE IV
COMPARISON BETWEEN MODELS ON SCFACE

| Method | Accuracy | Method | Accuracy |
|---|---|---|---|
| SSR [46] | 12.78% | LRFRW[9] | 24.30% |
| CSCDN[44] | 13.18% | DAlign [29] | 41.04% |
| CCA [47] | 15.11% | SKD [10] | 48.33% |
| C-RSDA[30] | 18.72% | SSPP-DAN [24] | 50.50% |
| DCA [31] | 17.44% | **Ours*** | **52.25%** |

### D. Evaluation on YouTubeFaces

We perform the evaluation on a large-scale YTF dataset [28] to demonstrate the generalization of the proposed method. YTF dataset consists of 3,425 videos of 1,595 subjects, including a total of 620 k images. We resize the first frame of each video $224 \times 224$ for the HR domain while downsampling the remaining frame to $16 \times 16$ for the LR domain. Among the downsampled LR domain samples, we divide them in half set and use each set for the train and test data respectively. We compare the proposed adaptation method with the 'Source Only' baseline under the same experimental settings of EK-LFH and SCface datasets. We confirm that the 'Proposed Adaptation' achieves 19.2%, which is much higher than 'Source Only' with 8.0% accuracy. This result supports the generalization performance of the proposed method in a large-scale dataset.

## V. CONCLUSION

A combined LRFR model with generative face augmentation (GFA) and spatial resolution adaptation (SRA) is proposed to reduce the domain differences at feature-level as well as image-level. In contrast to conventional methods, the proposed method does not use any labels on LR probe images to account for real-world face recognition. For image-level domain transfer, we first apply the proposed attention-guided GFA to the source HR images. Image-level domain transferred images are then trained in SRA to generate domain-invariant and discriminant feature distributions. Our unsupervised method of jointly training GFA and SRA outperforms the state-of-the-art supervised methods on the public dataset.

## References


[1] O. M. Parkhi *et al.*, "Deep face recognition," in *Proc. British Mach. Vision Conf.*, vol. 1, no. 3, 2015, p. 6.

[2] S. Hong, W. Im, J. Park, and H. Yang, "Deep cnn-based person identification using facial and clothing features," in *Proc. Summer Conf. Inst. Electron. Inf. Engineers*, 2016, pp. 2204–2207.

[3] Y. Zhong, J. Chen, and B. Huang, "Toward end-to-end face recognition through alignment learning," *IEEE Signal Process. Lett.*, vol. 24, no. 8, pp. 1213–1217, Aug. 2017.

[4] S. Gao, Y. Zhang, K. Jia, J. Lu, and Y. Zhang, "Single sample face recognition via learning deep supervised autoencoders," *IEEE Trans. Inf. Forensics Secur.*, vol. 10, no. 10, pp. 2108–2118, Oct. 2015.

[5] S. Gao, K. Jia, L. Zhuang, and Y. Ma, "Neither global nor local: Regularized patch-based representation for single sample per person face recognition," *Int. J. Comput. Vision*, vol. 111, no. 3, pp. 365–383, 2015.

[6] O. Abdollahi Aghdam, B. Bozorgtabar, H. Kemal Ekenel, and J.-P. Thiran, "Exploring factors for improving low resolution face recognition," in *Proc. Comput. Vision Pattern Recognit. Workshops*, 2019.

[7] Z. Cheng, X. Zhu, and S. Gong, "Surveillance face recognition challenge," 2018, *arXiv:1804.09691*.

[8] P. Li, L. Prieto, D. Mery, and P. Flynn, "Face recognition in low quality images: A survey," 2018, *arXiv:1805.11519*.

[9] 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.

[10] 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.

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

[12] B. F. Klare and A. K. Jain, "Heterogeneous face recognition using kernel prototype similarities," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 6, pp. 1410–1422, Jun. 2013.

[13] S. Liao, D. Yi, Z. Lei, R. Qin, and S. Z. Li, "Heterogeneous face recognition from local structures of normalized appearance," in *Proc. Int. Conf. Biometrics*, 2009, pp. 209–218.

[14] B. Li, H. Chang, S. Shan, and X. Chen, "Low-resolution face recognition via coupled locality preserving mappings," *IEEE Signal Process. Lett.*, vol. 17, no. 1, pp. 20–23, Jan. 2010.

[15] P. Li, J. Brogan, and P. J. Flynn, "Toward facial re-identification: Experiments with data from an operational surveillance camera plant," in *Proc. IEEE Int. Conf. Biometrics Theory, Appl. Syst.*, 2016, pp. 1–8.

[16] P. Li, M. L. Prieto, P. J. Flynn, and D. Mery, "Learning face similarity for re-identification from real surveillance video: A deep metric solution," in *Proc. IEEE Int. Joint Conf. Biometrics*, 2017, pp. 243–252.

[17] 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.

[18] M. Jian and K.-M. Lam, "Simultaneous hallucination and recognition of low-resolution faces based on singular value decomposition," *ACM Trans. Circuits Syst. Video Technol.*, vol. 25, no. 11, pp. 1761–1772, 2015.

[19] M.-C. Yang, C.-P. Wei, Y.-R. Yeh, and Y.-C. F. Wang, "Recognition at a long distance: Very low resolution face recognition and hallucination," in *Proc. IEEE Int. Conf. Biometrics*, 2015, pp. 237–242.

[20] S. Kolouri and G. K. Rohde, "Transport-based single frame super resolution of very low resolution face images," in *Proc. of Comput. Vision Pattern Recognit.*, 2015, pp. 4876–4884.

[21] C.-X. Ren, D.-Q. 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.

[22] S. Shekhar, V. M. Patel, and R. Chellappa, "Synthesis-based recognition of low resolution faces," in *Proc. IEEE Int. Joint Conf. Biometrics*, 2011, pp. 1–6.

[23] Q. Qiu, J. Ni, and R. Chellappa, "Dictionary-based domain adaptation methods for the re-identification of faces," in *Proc. Person Re-Identification*, 2014, pp. 269–285.

[24] S. Hong, W. Im, J. Ryu, and H. S. Yang, "Sspp-dan: Deep domain adaptation network for face recognition with single sample per person," in *Proc. IEEE Int. Conf. Image Process.*, 2017, pp. 825–829.

[25] M. Arjovsky, S. Chintala, and L. Bottou, "Wasserstein generative adversarial networks," in *Proc. Int. Conf. Mach. Learn.*, 2017, pp. 214–223.

[26] I. Goodfellow, "Nips 2016 tutorial: Generative adversarial networks," 2016, *arXiv:1701.00160*.

[27] M. Grgic, K. Delac, and S. Grgic, "Scface–surveillance cameras face database," *Multimedia Tools Appl.*, vol. 51, no. 3, pp. 863–879, 2011.

[28] L. Wolf, T. Hassner, and I. Maoz, "Face recognition in unconstrained videos with matched background similarity," in *Proc. Comput. Vision Pattern Recognit.*, 2011, pp. 529–534.

[29] S. P. Mudunuri, S. Venkataramanan, and S. Biswas, "Dictionary alignment with re-ranking for low-resolution nir-vis face recognition," *IEEE Trans. Inf. Forensics Secur.*, vol. 14, no. 4, pp. 886–896, Apr. 2019.

[30] 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, 2017.

[31] M. Haghighat and M. Abdel-Mottaleb, "Low resolution face recognition in surveillance systems using discriminant correlation analysis," in *Proc. IEEE Int. Conf. Autom. Face Gesture Recognit.*, 2017, pp. 912–917.

[32] X. Yu, B. Fernando, R. Hartley, and F. Porikli, "Super-resolving very low-resolution face images with supplementary attributes," in *Proc. Comput. Vision Pattern Recognit.*, 2018, pp. 908–917.

[33] X. Xie, Z. Cao, Y. Xiao, M. Zhu, and H. Lu, "Blurred image recognition using domain adaptation," in *Proc. IEEE Int. Conf. Image Process.*, 2015, pp. 532–536.

[34] S. Banerjee and S. Das, "Domain adaptation with soft-margin multiple feature-kernel learning beats deep learning for surveillance face recognition," 2016, *arXiv:1610.01374*.

[35] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *Proc. Comput. Vision Pattern Recognit.*, 2017, pp. 1125–1134.

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

[37] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Proc. Int. Conf. Comput. Vision*, 2017, pp. 2223–2232.

[38] Y. A. Mejjati, C. Richardt, J. Tompkin, D. Cosker, and K. I. Kim, "Unsupervised attention-guided image-to-image translation," in *Proc. Neural Inf. Process. Syst.*, 2018, pp. 3693–3703.

[39] S. Woo, J. Park, J.-Y. Lee, and I. So Kweon, "Cbam: Convolutional block attention module," in *Proc. Eur. Conf. Comput. Vision*, 2018, pp. 3–19.

[40] F. Wang *et al.*, "Residual attention network for image classification," in *Proc. Comput. Vision Pattern Recognit.*, 2017, pp. 3156–3164.

[41] R. A. Horn, "The hadamard product," in *Proc. Symp. Appl. Math*, vol. 40, 1990, pp. 87–169.

[42] W. Im, S. Hong, S.-E. Yoon, and H. S. Yang, "Scale-varying triplet ranking with classification loss for facial age estimation," in *Proc. Asian Conf. Comput. Vision*. Springer, 2018, pp. 247–259.

[43] S. Datta, G. Sharma, and C. Jawahar, "Unsupervised learning of face representations," in *Proc. IEEE Int. Conf. Autom. Face Gesture Recognit.*, 2018, pp. 135–142.

[44] Z. Wang, D. Liu, J. Yang, W. Han, and T. Huang, "Deep networks for image super-resolution with sparse prior," in *Proc. Int. Conf. Comput. Vision*, 2015, pp. 370–378.

[45] Y. Ganin and V. Lempitsky, "Unsupervised domain adaptation by backpropagation," in *Proc. of Int. Conf. Mach. Learn.*, 2015, pp. 1180–1189.

[46] J. Yang, J. Wright, T. S. Huang, and Y. Ma, "Image super-resolution via sparse representation," *IEEE Trans. Image Process.*, vol. 19, no. 11, pp. 2861–2873, Nov. 2010.

[47] Z. Wang, W. Yang, and X. Ben, "Low-resolution degradation face recognition over long distance based on cca," *Neural Comput. Appl.*, vol. 26, no. 7, pp. 1645–1652, 2015.