End-to-End Generative Adversarial Face Hallucination Through Residual In Internal Dense Network

Xiang Wang
Dept. of Computer Science
Lakehead University
Thunder Bay, Canada
xwang150@lakeheadu.ca

Yimin Yang
Dept. of Computer Science
Lakehead University
Thunder Bay, Canada
yyang48@lakeheadu.ca

Qixiang Pang
School of Engineering Tech., Trades and Aviation
Confederation College
Thunder Bay, Canada
qpang@confederationcollege.ca

Qiang Tang
Dept. of Electrial and Computer Engineering
The University of British Columbia Vancouver
Vancouver, Canada
qiangt@ece.ubc.ca

Shan Du *
Dept. of Comp. Sci., Math, Physics & Statistics
The University of British Columbia Okanagan
Kelowna, Canada
shan.du@ubc.ca

***Abstract*—Face hallucination has been a highly attractive computer vision research topic in recent years. It is still a particularly challenging task since the human face has a complex and delicate structure. In this paper, we propose a novel network structure, namely end-to-end Generative Adversarial Face Hallucination through Residual in Internal Dense Network (GAFH-RIDN), to hallucinate an unaligned tiny (32×32 pixels) low-resolution face image to its 8× (256×256 pixels) high-resolution counterpart. We propose a new architecture called Residual in Internal Dense Block (RIDB) for the generator and exploit an improved discriminator, Relativistic average Discriminator (RaD). In GAFH-RIDN, the generator is used to generate visually pleasant hallucinated face images, while the improved discriminator aims to evaluate how much input images are realistic. With continual adversarial learning, GAFH-RIDN is able to hallucinate perceptually plausible face images. Extensive experiments on large face datasets demonstrate that the proposed method significantly outperforms other state-of-the-art methods.**

***Index Terms*—Face Hallucination, Computer Vision, Generative Adversarial Network, Hallucinated Face Images**

I. Introduction

Face Hallucination (FH), also known as Face Super-Resolution (FSR), is a domain-specific image Super-Resolution (SR) problem, which refers to hallucinate the High-Resolution (HR) face images from their Low-Resolution (LR) counterparts. It is a significant task in the face analysis field, which is of remarkable benefit to computer vision applications such as face surveillance [1] and recognition [2]. However, face hallucination is an ill-posed inverse problem and particularly challenging since the LR image may correspond to many HR candidate images and has lost many crucial facial structures and components [3]–[5]. In order to hallucinate high quality face images, many FH methods have been proposed. Generally, we can classify these approaches into two categories: traditional methods and deep learning-based methods.

Many traditional methods have been proposed to address face hallucination tasks [6]–[8]. Baker and Kanade [6] presented the image pyramid model to learn the best relationship between LR and HR patches, which can reconstruct high-frequency details of LR face images. In [7], Wang and Tang employed eigen-transformation to build a linear mapping between LR and HR face subspaces. By adopting relationship between particular facial components, Yang $et\ al.$ [8] combined the face priors to recover facial information from HR image components.

Recently, deep learning-based methods have emerged and achieved the state-of-the-art performance [9]–[12]. Dong $et\ al.$ [9] firstly introduced deep learning-based SR method named SRCNN that directly learned an end-to-end mapping between HR images and LR images. In [10], Zhou $et\ al.$ presented the novel Bi-channel convolutional network to hallucinate face images in the wild. The Cascaded Bi-Networks (CBN) was presented by Zhu $et\ al.$ [11], in which two sub-networks (face hallucination and dense correspondence field estimation) were optimized alternately.

The limitation of the above face hallucination methods is that they utilize reconstruction loss such as $L1$ or $L2$ to optimize the hallucination process, which is prone to producing over-smoothed hallucinated images even though these models obtained higher Peak Signal-to-Noise Ratio (PSNR) value [13]. To address this problem, several Generative Adversarial Network (GAN) -based models were proposed [3], [4], [14]–[17]. It is proved that GAN-based models using powerful constraint losses are able to further generate visually realistic HR images [18]. Christian $et\ al.$'s work [14] extended GAN to the SR field and proposed an effective method, called SRGAN utilizing an adversarial loss and the perceptual loss. Following SRGAN, Wang $et\ al.$ [15] presented the ESRGAN by proposing new generator architecture and using improved perceptual loss. Yu and Porikli [19] proposed MTDN based

* Corresponding author: Shan Du (shan.du@ubc.ca). This work was supported by The University of British Columbia Okanagan [GR017752] and Lakehead University [11-50-16112406].

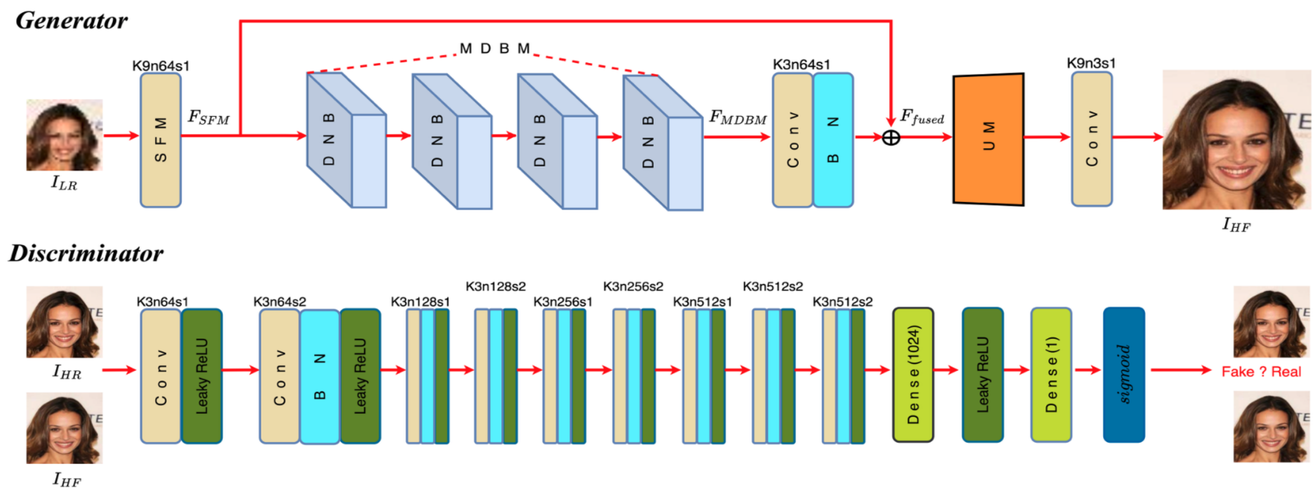

Fig. 1. The architecture of our end-to-end Generative Adversarial Face Hallucination through Residual in Internal Dense Network (GAFH-RIDN). I_{HF} represents HF image. I_{HR} and I_{LR} denote HR and LR face image respectively. K, n, and s represent kernel size, the number of feature maps and strides respectively. SFM is the Shallow Feature Module. MDBM describes the Multi-level Dense Block Module. UM is the Upsampling Module. DNB represents the Dense Nested Block as shown in Fig. 2. F_{SFM} denotes shallow feature maps. F_{MDBM} represents outputs of MDBM. By fusing F_{SFM} and F_{MDBM}, the fused feature F_{fused} is obtained.

on GAN. Nevertheless, when the input resolution is super low, it fails to recover high quality face images, leading to blurred patterns and severe artifacts.

The aforementioned GAN-based face hallucination models are prone to model collapse [3], [14]–[16], resulting in ghosting artifacts in the hallucinated results, especially when the input image resolution is extremely low. To address this problem, in this paper, we propose a novel GAN-based FH method, end-to-end Generative Adversarial Face Hallucination through Residual in Internal Dense Network (GAFH-RIDN), as shown in Fig. 1. The contributions of this paper are mainly in four aspects:

1) Our proposed method is capable of hallucinating an LR (32×32 pixels) unaligned tiny face image to a Hallucinated Face (HF) image (256×256 pixels) with an ultra upscaling factor 8×.

2) We propose the Residual in Internal Dense Block (RIDB), which boosts the flow of features through the generator, alleviates the gradient vanishing problem, and provides hierarchical features for the hallucination process.

3) We exploit the Relativistic average Discriminator (RaD) [20], which evaluates the probability that the given HR face images are more realistic than HF images.

4) Contrary to classical face hallucination methods [3], [4], [21], our method does not involve any prior information or claim facial landmark points for its hallucinating, which facilitates the whole training process and enhances the model robustness.

II. PROPOSED METHOD

In this section, we will first describe the proposed architecture and demonstrate the Residual in Internal Dense Block (RIDB). Next, we will discuss the improved discriminator. Finally, we will present the perceptual and adversarial losses function used in the GAFH-RIDN. The architecture of GAFH-RIDN is shown in Fig. 1.

A. Network Architecture

As shown at the top of Fig. 1, the proposed generator mainly consists of three stages: Shallow Feature Module (SFM), Multi-level Dense Block Module (MDBM), and Upsampling Module (UM). The LR face image I_{LR} is fed into the SFM as the initial input. At the end, the hallucinated face image I_{HF} is obtained from the UM. As for the SFM, we utilize one convolutional (Conv) layer to extract the shallow feature maps. It can be expressed as follows:

$$F_{SFM} = f_{Conv}(I_{LR}) \tag{1}$$

where f_{Conv} represents the Conv operation in the SFM. F_{SFM} denotes the shallow (low-level) features and serves as the input to the MDBM. The following module MDBM is built up by multiple Dense Nested Blocks (DNB) formed by several RIDBs, which will be discussed in the next subsection. The procedure of high-level feature extraction in MDBM can be formulated as:

$$F_{MDBM} = f_{DNB,i}(f_{DNB,i-1}(\cdots(f_{DNB,1}(F_{SFM}))\cdots)) \tag{2}$$

where $f_{DNB,i}$ denotes high-level feature extraction of the i-th DNB, F_{MDBM} represents the high-level feature extracted by MDBM. As for each DNB, it includes 3 RIDBs cascaded by residual connections and one scale layer, as shown in Fig. 2. It can be formulated as:

$$F_{DNB,i} = \alpha F_{i,j}(F_{i,j-1}(\cdots F_{i,1}(F_{DNB,i-1})\cdots)) + F_{DNB,i-1} \tag{3}$$

where $F_{DNB,i-1}$, $F_{DNB,i}$ denotes the input and output of i-th DNB, $F_{i,j}$ represents the j-th RIDB of the i-th DNB. We assign α to be 0.2 in the scale layer. Next, the low-level and high-level features should be fused to boost hallucination performance via skip connection. Let F_{fused} denotes the fused feature, the feature fusion process can be expressed as:

$$F_{fused} = f_{Conv}(F_{MDBM}) + F_{SFM} \tag{4}$$

Furthermore, the fused feature F_{fused} is passed to the UM followed by one Conv layer. And then, the fused feature is transformed from the LR space to the HR space through

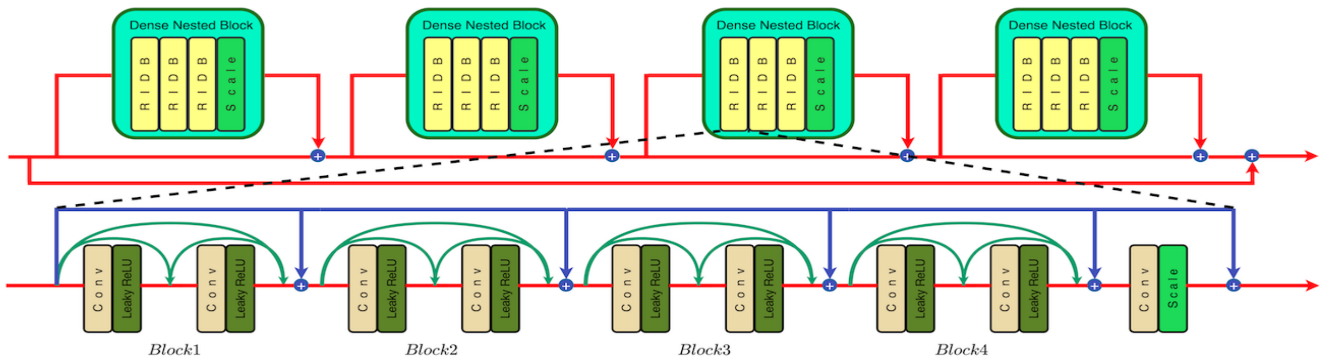

Fig. 2. **Top**: Dense Nested Block (DNB) composed of multiple RIDBs. **Bottom**: The architecture of our proposed Residual in Internal Dense Block (RIDB).

upsampling layers in the UM. The hallucination process can be formulated as:

$$I_{HF} = f_{UM}(F_{fused}) = H_{GAFH-RIDN}(I_{LR}) \quad (5)$$

where f_{UM} represents the upsampling operation in the UM, $H_{GAFH-RIDN}$ denotes the function of our GAFH-RIDN. Finally, we obtain the HF image I_{HF}.

B. Residual in Internal Dense Block

As mentioned in Sec.1, we propose a novel architecture RIDB for the generator, which is used to form the DNB (as shown in Fig. 2). The proposed RIDB is able to extract hierarchical features and address the vanishing-gradient problem, which is the commonly encountered issue in [14]–[16], [22], [23]. The proposed RIDB is made up of four internal dense blocks and all the internal dense blocks are cascaded through residual connections performing identity mapping. The architecture of the RIDB is expressed as:

$$F_{RIDB,p} = F_{p,q}(F_{p,q-1}(\cdot\cdot F_{p,1}(F_{RIDB,p-1})\cdot\cdot)) + F_{RIDB,p-1} \quad (6)$$

where $F_{RIDB,p-1}$ and $F_{RIDB,p}$ denote the input and output of the p-th RIDB respectively, $F_{p,q}$ represents the q-th internal dense block of p-th RIDB. In addition, an internal dense block is a composition of two groups of the Conv layer followed by the LeakyReLU activation layer. And the two groups are linked by dense skip connections. Each internal dense block can be calculated as follows:

$$F_{q,k} = \delta(W_{q,k}[F_{q,k=1}, F_{q,k=2}]) \quad (7)$$

where $F_{q,k}$ represents the output of k-th Conv layer of q-th internal dense block. $[F_{q,k=1}, F_{q,k=2}]$ refers to the concatenation of feature maps in q-th internal dense block. $W_{q,k}$ is the weights of the k-th Conv layer. δ denotes the LeakyReLU activation. By involving residual learning and more dense connections in the RIDB, the feature maps of each layer are propagated into all succeeding layers, promoting an effective way for the generator to extract hierarchical features and strengthening the flow of graidents through the network. Thus, our proposed method is capable of obtaining abundant hierarchical feature information and alleviating the vanishing-gradient problem.

C. Improved Discriminator

Instead of using the discriminator of Standard GAN (SGAN) [26], inspired by [20], we adopt the Relativistic average Discriminator (RaD) in our method. Thanks to RaD, the discriminator of GAFH-RIDN has the ability to distinguish how the given HR face image is more authentic than the hallucinated face image. The architecture of our discriminator is shown at the bottom of Fig. 1. The limitation of the SGAN in [14], [16], [26] is that they only concentrate on increasing the probability that fake samples belong to real rather than decreasing the probability that real samples belong to real simultaneously. In other words, the standard discriminator ignores real samples during the learning procedure [20]. As a result, the model can not provide sufficient gradients when updating the generator, which causes the problem of gradient vanishing for training the generator. The standard discriminator can be expressed as:

$$D(x) = \sigma(C(x)) \quad (8)$$

where x can be either I_{HR} or I_{HF} in this context, σ represents the sigmoid function, and $C(x)$ denotes the output of non-transformed discriminator. As Eq. 8 shows, the standard discriminator only evaluates the probability for a given real sample or a generated sample. According to [20], RaD takes into consideration that how a given real sample is more authentic than a given generated sample. The RaD can be formulated as:

$$D(x_r, x_f) = \sigma(C(x_r) - E_{x_f}[C(x_f)]) \quad (9)$$

where E_{x_f} denotes the average of the fake samples in one batch. Contrary to the standard discriminator, as Eq. 9 shows, the probability predicted by RaD relies on both real sample x_r and fake sample x_f, which is capable of making the discriminator to become relativistic. In our GAFH-RIDN, we can optimize the RaD by L_D^{adv} based on Eq. 10, and the generator is updated by L_G^{adv}, as in Eq. 11.

$$\begin{aligned} L_D^{adv} = &- \mathbb{E}_{I_{HR}} \sim p_{(I_{HR})}\left[log\left(D(I_{HR}, I_{HF})\right)\right] \\ &- \mathbb{E}_{I_{HF}} \sim p_{(I_{HF})}\left[log\left(1 - D\left(I_{HF}, I_{HR}\right)\right)\right] \end{aligned} \quad (10)$$

$$\begin{aligned} L_G^{adv} = &- \mathbb{E}_{I_{HR}} \sim p_{(I_{HR})}\left[log\left(1 - D(I_{HR}, I_{HF})\right)\right] \\ &- \mathbb{E}_{I_{HF}} \sim p_{(I_{HF})}\left[log\left(D\left(I_{HF}, I_{HR}\right)\right)\right] \end{aligned} \quad (11)$$

where I_{HR} and I_{HF} denote HR images and HF images respectively, $D(\cdot)$ describes the probability predicted by RaD, $\mathbb{E}$ represents the expectation, $I_{HR} \sim P_{I_{HR}}$ and $I_{HF} \sim P_{I_{HF}}$ represents the HR images distribution and HF images distribution respectively. Because of this property, our proposed GAFH-RIDN is capable of allowing the probability of I_{HR} being real to decrease while letting the probability of I_{HF} being real increase and benefiting from gradients of both I_{HR} and I_{HF} in the adversarial training. Therefore the proposed method can address the gradient vanishing problem. Our discriminator contains 9 Conv layers with the number

Fig. 3. Comparison of visual results with state-of-the-art methods on scaling factor 8×. (a) HR images, (b) LR inputs, (c) Bicubic interpolation, (d) Results of SRGAN [14], (e) Results of ESRGAN [15], and (f) Our results

TABLE I
QUANTITATIVE COMPARISON ON CELEBA DATASET FOR SCALING FACTOR 8×, IN TERMS OF AVERAGE PSNR(DB) AND SSIM. NUMBERS IN BOLD ARE THE BEST EVALUATION RESULTS AMONG STATE-OF-THE-ART METHODS.

Method	Bicubic	VDSR [12]	CBN [11]	SRGAN [14]	FSRFCH [24]	TDN [5]	Kim et al. [25]	ESRGAN [15]	Ours
PSNR	22.90	19.58	18.77	20.64	23.14	22.66	22.96	20.32	**24.28**
SSIM	0.65	0.57	0.54	0.62	0.68	0.66	0.69	0.57	**0.71**

of 3x3 kernels and the stride of 1 or 2 alternately. The channels of feature maps increase by a factor 2, from 64 to 512. The resulting 512 feature maps are passed through two dense layers. Finally, after the sigmoid activation layer, RaD estimates the probability that the given HR face images are more realistic than HF images.

D. Perceptual Loss

Taking advantage of perceptual loss $L_{perceptual}$ is able to promote ulteriorly detail enhancement [27], [28]. We adopt the pre-trained VGG-19 [29] as the feature extractor to obtain feature representation used to calculate $L_{perceptual}$. We extract low-level feature maps of HR and HF images obtained by the 3^{rd} Conv layer before the 4^{th} maxpooling layer respectively. HR and HF feature maps are defined as $\phi_{3,4}$. $L_{perceptual}$ is defined as follows:

$$L_{perceptual} = \frac{1}{WH}\sum_{h=1}^{H}\sum_{w=1}^{W}\|\phi_{3,4}(I_{HR}) - \phi_{3,4}(I_{HF})\|^2 \tag{12}$$

where ϕ represents the feature extractor and W, H denote the dimensions of feature maps.

E. Total Loss

The total loss function L_{total} for generator can be represented as two parts: $L_{perceptual}$ and L_G^{adv}. We introduce the perceptual loss to enhance perceptual quality of the HF image from the visual aspect. In addition, an adversarial loss is expected to improve the fidelity of the HF image. The formula is defined as follows:

$$L_{total} = \alpha L_{perceptual} + \beta L_G^{adv} \tag{13}$$

where α and β are corresponding hyper-parameters used to balance $L_{perceptual}$ and L_G^{adv}. We empirically set α = 1, β = 10^{-3} respectively.

III. EXPERIMENTS

In this section, we will first present the details of datasets and implementation. Next, we will discuss the comparisons with the state-of-the-art methods [5], [11], [12], [14], [15], [24], [25] qualitatively and quantitatively.

A. Implementation Details

We conducted experiments on the large-scale face image dataset, CelebA [30]. It consists of 202,599 face images of 10,177 celebrities. We randomly selected 162,048 HR face images as the training set, and the next 40,511 images were used as the testing set. We cropped the HR face images and resized them to 256×256 pixels, and then obtained LR (32×32 pixels) input images by downsampling HR images using bicubic interpolation with a downsampling factor of 8×. In the proposed generator, we set the number of DNBs to

4, and totally 12 RIDBs were used. In the training phase, we trained the proposed method for 10000 epochs and the training batch size was set to 8. We used Adam with β_1 = 0.9, β_2 = 0.999 to optimize the proposed method. The learning rate was set to 10^{-4}. We alternately updated the generator and discriminator.

B. Comparisons with State-of-the-art Methods

In the experiments, we compared the proposed method with the state-of-the-art methods [5], [11], [12], [14], [15], [24] qualitatively and quantitatively.

Qualitative Comparison: Qualitative results among these methods are shown in Fig. 3. We observe that the bicubic interpolation produces heavy blur and fails to generate clear textures. For SRGAN [14], it outputs noticeable artifacts around facial components, especially in the eyes, nose, and mouth regions. In particular, ESRGAN [15] produces unrealistic textures and involves severe ghosting artifacts. In contrast, it is obvious that our proposed method is capable of producing visually pleasant and authentic HF images.

Quantitative Comparison: Table 1 shows the quantitative comparison on 8× HF images. The results demonstrate that our proposed method achieves the best performance among all methods. Especially, our method produces the highest score of 24.28dB/0.71 in terms of PSNR and SSIM respectively. Furthermore, compared with the second-best FSRFCH [24] 23.14dB/0.68, our method outperforms it with a large margin of 1.14dB/0.03. The performance proves the effectiveness of the proposed RIDB and the optimized RaD used in our method.

IV. CONCLUSIONS

In this paper, we proposed a novel end-to-end face hallucination method (GAFH-RIDN) to hallucinate a tiny LR (32×32 pixels) unaligned face image to its 8× HR (256×256 pixels) version. By exploiting Residual in Internal Dense Block (RIDB) and Relativistic average Discriminator (RaD), our method successfully produced photo-realistic hallucinated face images. Extensive experiments demonstrated that GAFH-RIDN was superior to the state-of-the-art methods on the face benchmark qualitatively and quantitatively.

REFERENCES

[1] W. W. Zou and P. C. Yuen, "Very low resolution face recognition problem," *IEEE Transactions on image processing*, vol. 21, no. 1, pp. 327–340, 2011.

[2] J. Yang, L. Luo, J. Qian, Y. Tai, F. Zhang, and Y. Xu, "Nuclear norm based matrix regression with applications to face recognition with occlusion and illumination changes," *IEEE transactions on pattern analysis and machine intelligence*, vol. 39, no. 1, pp. 156–171, 2016.

[3] Y. Chen, Y. Tai, X. Liu, C. Shen, and J. Yang, "Fsrnet: End-to-end learning face super-resolution with facial priors," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 2492–2501.

[4] X. Yu, B. Fernando, R. Hartley, and F. Porikli, "Super-resolving very low-resolution face images with supplementary attributes," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 908–917.

[5] X. Yu and F. Porikli, "Face hallucination with tiny unaligned images by transformative discriminative neural networks," in *Thirty-First AAAI conference on artificial intelligence*, 2017.

[6] S. Baker and T. Kanade, "Limits on super-resolution and how to break them," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 9, pp. 1167–1183, 2002.

[8] C.-Y. Yang, S. Liu, and M.-H. Yang, "Structured face hallucination," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2013, pp. 1099–1106.

[7] X. Wang and X. Tang, "Hallucinating face by eigentransformation," *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 35, no. 3, pp. 425–434, 2005.

[9] C. Dong, C. C. Loy, K. He, and X. Tang, "Image super-resolution using deep convolutional networks," *IEEE transactions on pattern analysis and machine intelligence*, vol. 38, no. 2, pp. 295–307, 2015.

[10] E. Zhou, H. Fan, Z. Cao, Y. Jiang, and Q. Yin, "Learning face hallucination in the wild," in *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence*, 2015, pp. 3871–3877.

[11] S. Zhu, S. Liu, C. C. Loy, and X. Tang, "Deep cascaded bi-network for face hallucination," in *European conference on computer vision*. Springer, 2016, pp. 614–630.

[12] J. Kim, J. Kwon Lee, and K. 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.

[13] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE transactions on image processing*, vol. 13, no. 4, pp. 600–612, 2004.

[14] C. Ledig, L. Theis, F. Huszár, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang *et al.*, "Photo-realistic single image super-resolution using a generative adversarial network," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 4681–4690.

[15] X. Wang, K. Yu, S. Wu, J. Gu, Y. Liu, C. Dong, Y. Qiao, and C. Change Loy, "Esrgan: Enhanced super-resolution generative adversarial networks," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018, pp. 0–0.

[16] X. Yu and F. Porikli, "Ultra-resolving face images by discriminative generative networks," in *European conference on computer vision*. Springer, 2016, pp. 318–333.

[17] ——, "Hallucinating very low-resolution unaligned and noisy face images by transformative discriminative autoencoders," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 3760–3768.

[18] R. Zhang, P. Isola, A. A. Efros, E. Shechtman, and O. Wang, "The unreasonable effectiveness of deep features as a perceptual metric," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 586–595.

[19] X. Yu, F. Porikli, B. Fernando, and R. Hartley, "Hallucinating unaligned face images by multiscale transformative discriminative networks," *International Journal of Computer Vision*, vol. 128, no. 2, pp. 500–526, 2020.

[20] A. Jolicoeur-Martineau, "The relativistic discriminator: a key element missing from standard gan," *arXiv preprint arXiv:1807.00734*, 2018.

[21] X. Ma, J. Zhang, and C. Qi, "Hallucinating face by position-patch," *Pattern Recognition*, vol. 43, no. 6, pp. 2224–2236, 2010.

[22] T. Tong, G. Li, X. Liu, and Q. Gao, "Image super-resolution using dense skip connections," in *Proceedings of the IEEE International Conference on Computer Vision*, 2017, pp. 4799–4807.

[23] Y. Zhang, Y. Tian, Y. Kong, B. Zhong, and Y. Fu, "Residual dense network for image super-resolution," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 2472–2481.

[24] X. Yu, B. Fernando, B. Ghanem, F. Porikli, and R. Hartley, "Face super-resolution guided by facial component heatmaps," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018, pp. 217–233.

[25] D. Kim, M. Kim, G. Kwon, and D.-S. Kim, "Progressive face super-resolution via attention to facial landmark," *arXiv preprint arXiv:1908.08239*, 2019.

[26] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *Advances in neural information processing systems*, 2014, pp. 2672–2680.

[27] J. Johnson, A. Alahi, and L. Fei-Fei, "Perceptual losses for real-time style transfer and super-resolution," in *European conference on computer vision*. Springer, 2016, pp. 694–711.

[28] A. Dosovitskiy and T. Brox, "Generating images with perceptual similarity metrics based on deep networks," in *Advances in neural information processing systems*, 2016, pp. 658–666.

[29] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[30] Z. Liu, P. Luo, X. Wang, and X. Tang, "Deep learning face attributes in the wild," in *Proceedings of the IEEE international conference on computer vision*, 2015, pp. 3730–3738.