We are IntechOpen,
the world's leading publisher of
Open Access books
Built by scientists, for scientists

6,900
Open access books available

185,000
International authors and editors

200M
Downloads

154
Countries delivered to

Our authors are among the
TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities

CLARIVATE ANALYTICS
BOOK CITATION INDEX
INDEXED

WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com

Chapter

Style-Based Unsupervised Learning for Real-World Face Image Super-Resolution

Ahmed Cheikh Sidiya and Xin Li

Abstract

Face image synthesis has advanced rapidly in recent years. However, similar success has not been witnessed in related areas such as face single image super-resolution (SISR). The performance of SISR on real-world low-quality face images remains unsatisfactory. In this paper, we demonstrate how to advance the state-of-the-art in face SISR by leveraging style-based generator in unsupervised settings. For real-world low-resolution (LR) face images, we propose a novel unsupervised learning approach by combining style-based generator with relativistic discriminator. With a carefully designed training strategy, we demonstrate our converges faster and better suppresses artifacts than Bulat's approach. When trained on an ensemble of high-quality datasets (CelebA, AFLW, LS3D-W, and VGGFace2), we report significant visual quality improvements over other competing methods especially for real-world low-quality face images such as those in Widerface. Additionally, we have verified that both our unsupervised approaches are capable of improving the matching performance of widely used face recognition systems such as OpenFace.

Keywords: single image super-resolution (SISR), unsupervised learning, degradation modeling, real-world face images

1. Introduction

With recent advancements in deep learning algorithms [1], Single Image Superresolution (SISR) has seen a significant advance in performance in terms of objective metrics like peak signal-to-noise-ratio (PSNR). With generative adversarial networks (GAN) such as improvements of objective quality metric have been extended to the visual quality of super-resolved images [2]. However, most of the existing deep learning algorithms for solving SISR problems are categorized as supervised; in that they rely on paired high-resolution (HR) and low-resolution (LR) images to optimize the neural network weights. The HR images are downsampled using algorithms (like bicubic downsampling) to create the corresponding LR ones. These artificially created LR data deviate significantly from the complex real word degradation model and with that a rapid decrease in performance is observed when neural networks trained on artificial LR that are tested on real-world LR images [3].

In this chapter, we will focus on solving the problem of image superresolution for face images. Super-resolving low resolution face images can help solve crucial tasks such as person identification and recognition in the real world. To make our solution work for real-world LR face images, we borrow ideas from recent advances in style transfer [4] and image synthesis [5]. Style transfer refers to the task of transforming one image from one style to another (e.g., photo to painting, daytime to nighttime, and summer to winter). An important motivation behind our approach is to treat SISR as a style transfer problem which does not require pairing the HR-LR training data. In our unsupervised learning approach, we only assume two uncorrelated datasets: one is a collection of real-world LR images and the other HR images.

In the next sections, we will first review some related works including convolutional neural network (CNN), generative adversarial networks (GAN), image synthesis, and style transfer. We will then present an unsupervised approach that works for real-world LR face images. The key idea is to combine style-based generator [5] with relativistic discriminator [6] within a recently developed cycle-consistent GAN (CycleGAN) framework [4]. We will show that both our approaches outperform previous state-of-the-art ones.

2. Related works

In this section, we will first present to the reader the convolution neural networks and go through the different types of functions used in such networks. Our focus will be on the main convolution operation. We will talk about the first paper that showed that convolution neural networks can outperform model-based approaches in the task of image super-resolution [7]. We will talk about generative adversarial networks (GAN) and show that adding a discriminator can significantly improve the visual quality of the superresolved image [2]. We will define image synthesis task and present the latest advancements in the field. Finally, we will talk about the style transfer problem and different architectures used to solve it; our focus will be on the most popular one: CycleGAN [4].

2.1 Convolutional neural networks

2.1.1 Definition

Neural networks are a class of machine learning algorithms that are modeled loosely on the mechanism of human brain (e.g., neocognitron [8]). It consists of thousands or even millions of simple processing units that are densely interconnected. Most existing neural networks are organized into layers of nodes (simple processing units). They usually feed-forward information from input data to the output in one direction. An individual node might be connected to several nodes in the layer beneath it from which it receives data and several nodes in the layer above it for which it sends data. **Figure 1** shows an example of feed-forward neural network. To each of its incoming connections, some nodes will assign a number called “weight,” multiply the input coming from the connection with its corresponding weights; other nodes will sum the results and add a value called bias. In other nodes, a non-linearity called activation function is included which models the biological firing of neurons in human brains [9].

The Convolutional Neural Network (CNN) are a type of neural networks that are often designed to work on two-dimensional data such as image signals. In CNN, the most basic operation is called “convolution” implementing a linear filter and

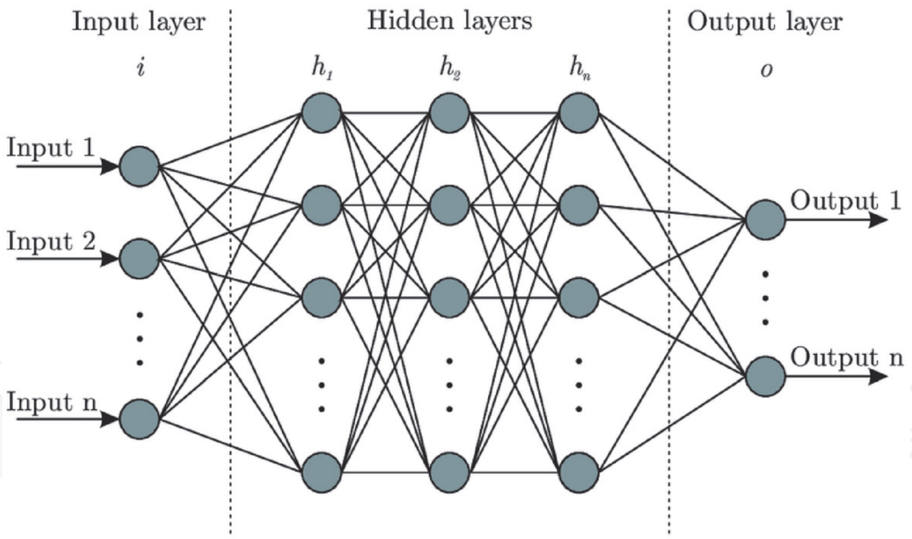


Figure 1.
Graphical representation of feed-forward neural network.

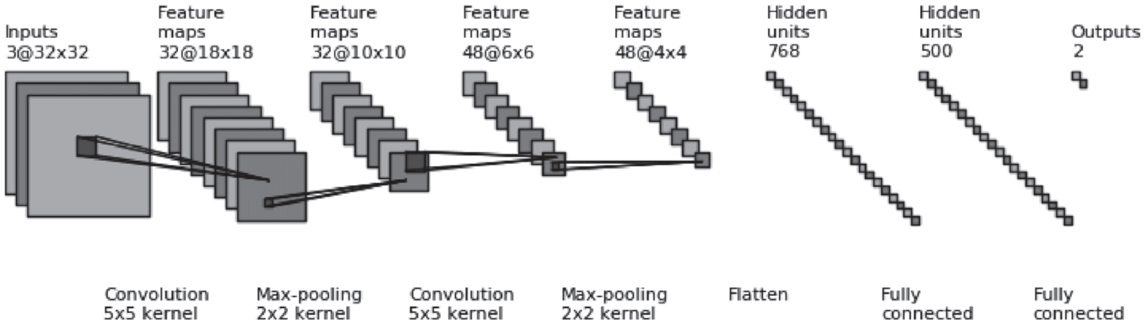


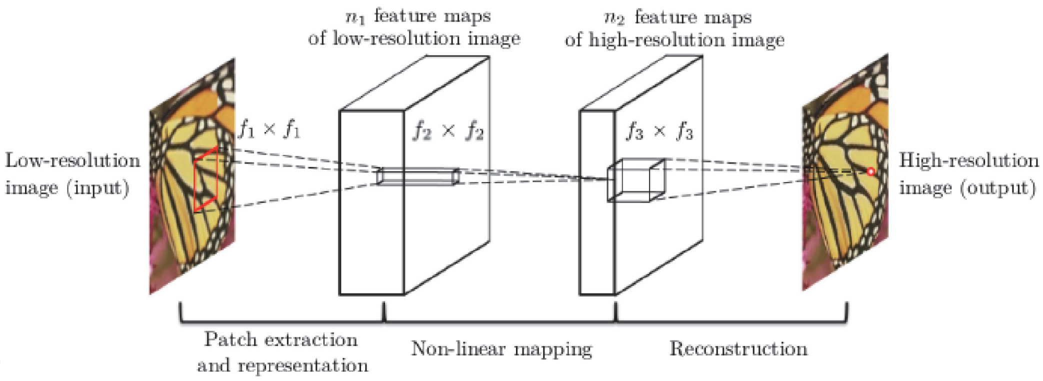
Figure 2.
Example of convolution neural network.

modeling simple cells in human brains [10]. In the context of convolutional neural networks, a convolution is a linear filtering operation that involves multiplying the input with the weights similar to the traditional neural network. Given the nature of 2D inputs, the multiplication is done between an array of data and 2D array of weights (often called filter or kernel).

The filter is smaller than the data and the multiplication operation between the filter and the data is the dot product. The dot product is the element-wise multiplication and summation resulting in one value. Having a filter or a kernel smaller than the input data enables the sliding of the kernel over the whole input, therefore giving the trained weights of the filter the ability to detect features anywhere in the image. Convolutional neural networks might also consist of max-pooling operations, used to down-sample the input (modeling complex cells in human brain [10]). **Figure 2** shows a graphical representation of neural network.

2.1.2 Image super-resolution using convolutional neural networks

In [7], the authors present the first convolutional neural network called SRCNN that outperforms traditional model-based approaches for the task of single image super-resolution (SISR). The key idea underlying SRCNN is to learn a nonlinear mapping from the space of LR images to that of HR ones. The work of SRCNN is under the framework of supervised learning with the assumption of paired training data (artificial LR images are generated by down-sampling of HR images). Their



IntechOpen

Figure 3.
Convolution neural network architecture for [7].

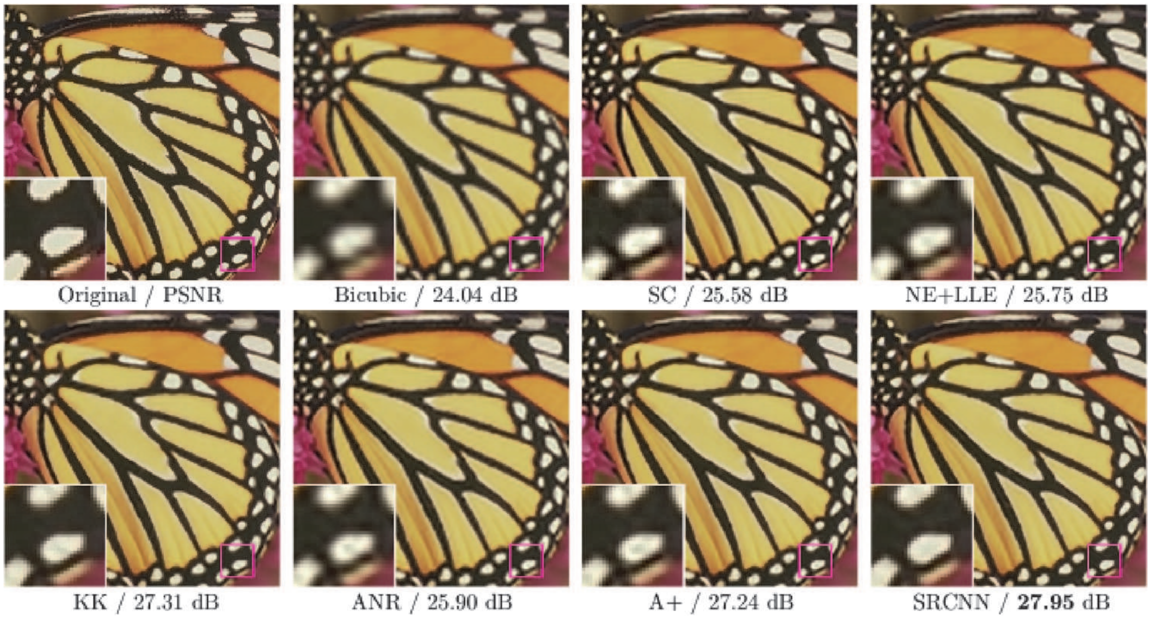


Figure 4.
Comparisons between [7] and state-of-the-art model based methods.

convolution neural network as shown in **Figure 3** consists of two layers. **Figure 4** shows the comparison between [7] and traditional model-based SR methods such as sparse coding [11].

With advancements in deep neural networks, deeper architectures powered by residual learning [12] has led to FSRCNN [13], DRCN [14], VDSR [15], EDSR [16] and LapSRN [17], RDN [18], and RCAN [19]. With deeper and densely connected networks, the performance of SISR has increased steadily at the price of higher computational complexity. With millions of parameters, EDSR and RCAN have advanced the state-of-the-art in supervised learning-based SISR. For face images, SISR has also been studied in recent works (e.g., Super-FAN [20] and FSRNet [21]).

2.2 Generative adversarial networks (GAN)

2.2.1 Definition

In [22], Ian Goodfellow presented a novel system for the task of data generation. This system is called generative adversarial network (GAN) consists of two interacting subnetworks (generator and discriminator) as shown in **Figure 5**. A generator subnetwork is responsible for generating synthetic data capturing the

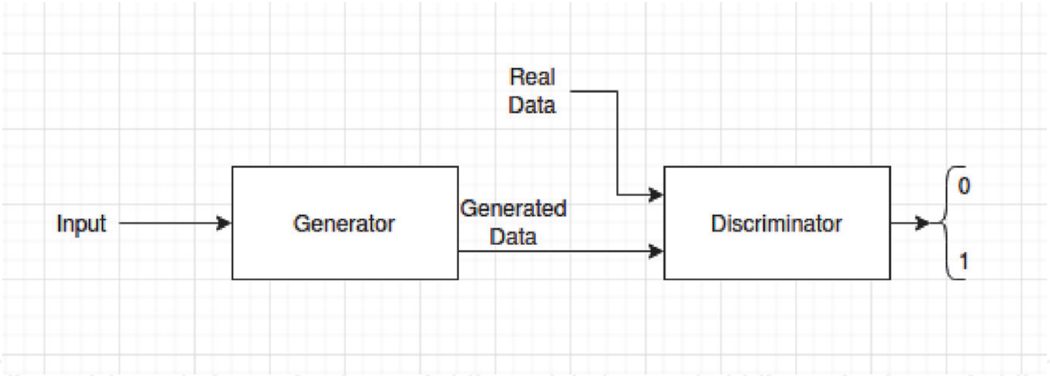


Figure 5.
Architecture of GANs.

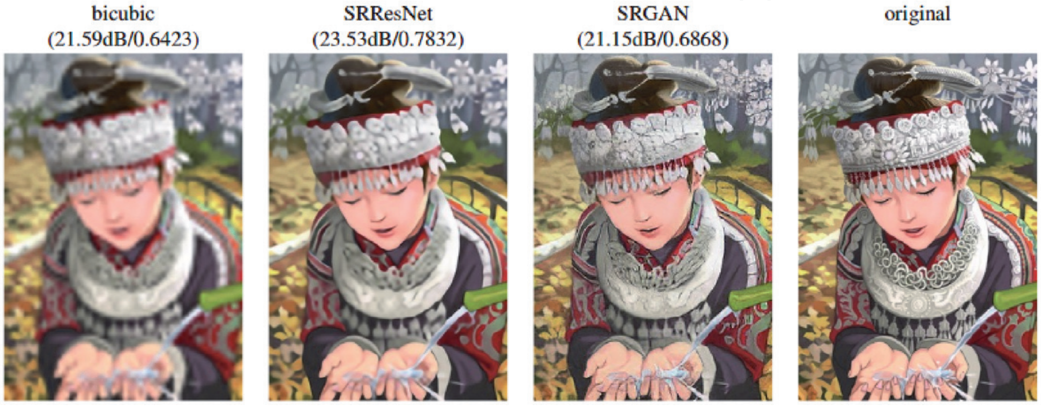


Figure 6.
Comparison between SRGAN and SRResnet.

data distribution and a discriminator subnetwork for estimating the probability that a sample comes from the real training data rather than synthetic. Through the interplay between two subnetworks, the generator and discriminator networks can be trained together by a minimax two-player game. The invention of GAN opens the door to construct a whole new class of powerful generative models which have found numerous applications in low-level vision including SISR, face image synthesis, and style transfer.

2.2.2 Single image super-resolution using generative adversarial networks

In SRGAN [2], the authors showed that using a GAN-based architecture for the task of single image super-resolution leads to noticeable improvements in terms of subjective visual quality despite the sacrifice on traditional objective quality metric such as PSNR. In the construction of SRGAN, residue network for SR-called SRResnet is used as the generator; a separated discriminator inspired by Deep Convolutional GAN (DCGAN) [23] is constructed to tell apart real SR from fake SR. **Figure 6** shows the visual quality improvement when using a GAN-based architecture compared to using a generator without a discriminator.

2.3 Face image synthesis

Another successful application of GAN [2] is to generate high-fidelity face images that do not even exist in the real world. Radford et al. designed a variation of GAN architecture called Deep Convolutional GAN (DCGAN) [23] to generate face images; however, their results suffered from noticeable artifacts in synthesized images. More recently, self-attention (SAGAN) [24] used an attention mechanism to help minimize undesirable artifacts in generated images. Cleverly, designing loss functions for both discriminator and generator has shown impressive improvements in terms of convergence and artifacts suppression for the GAN networks.

Before 2017, most generated faces were still of low resolution, with the highest resolution equal to 128 × 128. In [5, 25], it was shown that progressively training the generator network helped generate face images up to 1024 × 1024 resolution. **Figure 7** shows an example of visual quality improvements in face image synthesis in the past 5 years, for example, from Progressive GAN [25] to StyleGAN [5] and its enhancement version StyleGAN2 [26].

IntechOpen

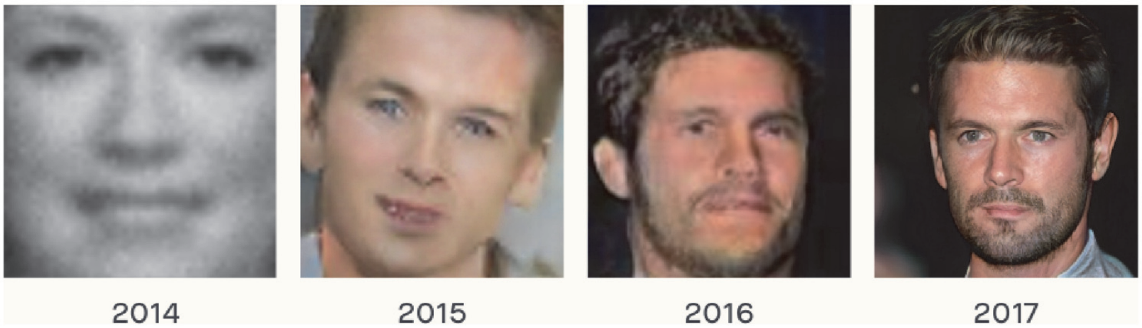


Figure 7.
Example of the progression made in GAN-based face synthesis from 2014 to 2017 (cited from [27]).

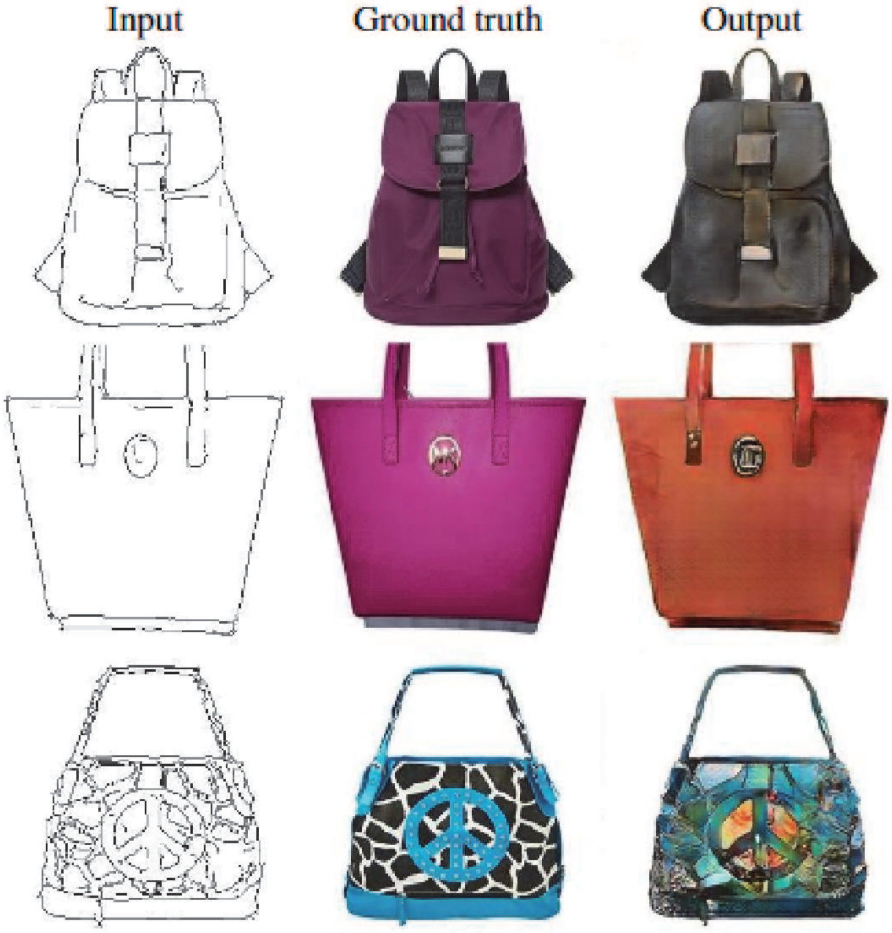


Figure 8.
Results from pix2pix [28].

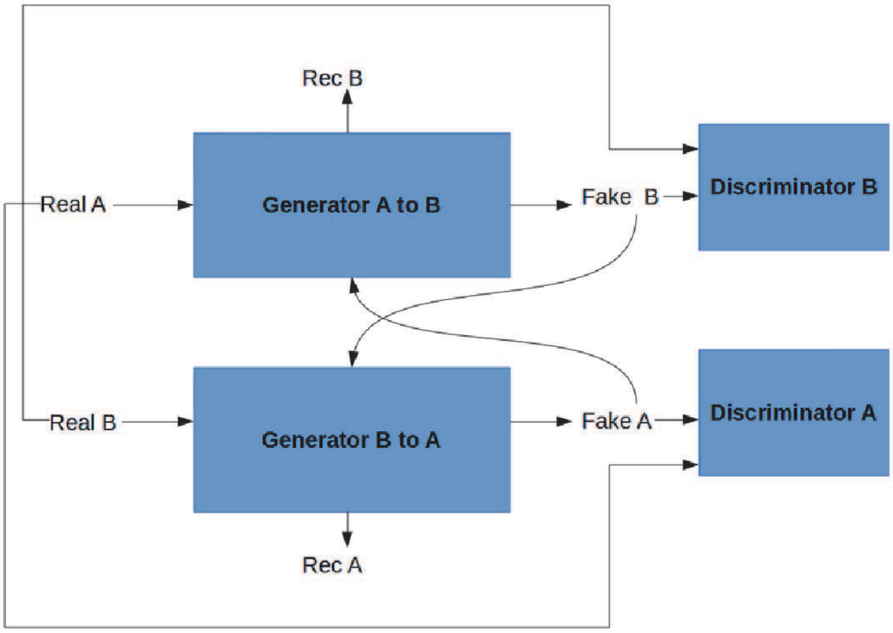


Figure 9.
CycleGAN [29] architecture.

2.4 Style transfer (CycleGAN)

Recently, GAN-based architectures were used for the task of style transfer, that is, translating one image from one style to another style (e.g., from sketch to real image). In [28], another GAN-based architecture called pix2pix was developed to transfer images from one style into another. **Figure 8** shows the example of translating sketch images of hand bags into real images. These works are based on an architecture called conditional GAN (cGAN) [30] in which the data instead of random noise are fed to (provided as the condition) both the generator and discriminator. However, pix2pix [28] is a supervised learning technique, and it requires the existence of groundtruth data.

To extend the style transfer to the domain where groundtruth is unavailable, an unsupervised architecture called cycle-consistent GAN (CycleGAN) was proposed in [29]. In CycleGAN [29], two parallel GAN architectures are trained concurrently: the first one to map from source domain to target domain and the second one to map from target domain back to source domain. The new insight brought by CycleGAN [29] is the enforcement of cycle-consistency, that is, when an image X is translated from source domain to target domain via forward mapping f and then translated back to the original domain via background mapping g, the result should approach the original image ($x \approx g(f(x))$). **Figure 9** shows an example of the CycleGAN [29] architecture with two generators and two discriminators.

3. Unsupervised approach

3.1 Overview of the method

We are interested in solving the problem of image face super-resolution for real world LR data. Unlike artificial LR data, the ground-truth is unavailable for real-world LR data. For such blind SR problem, we propose to tackle it as style transfer, that is, the transfer between LR and HR image data. We mainly focus on an asymmetrical formulation of style transfer problem in one direction: from LR to HR.

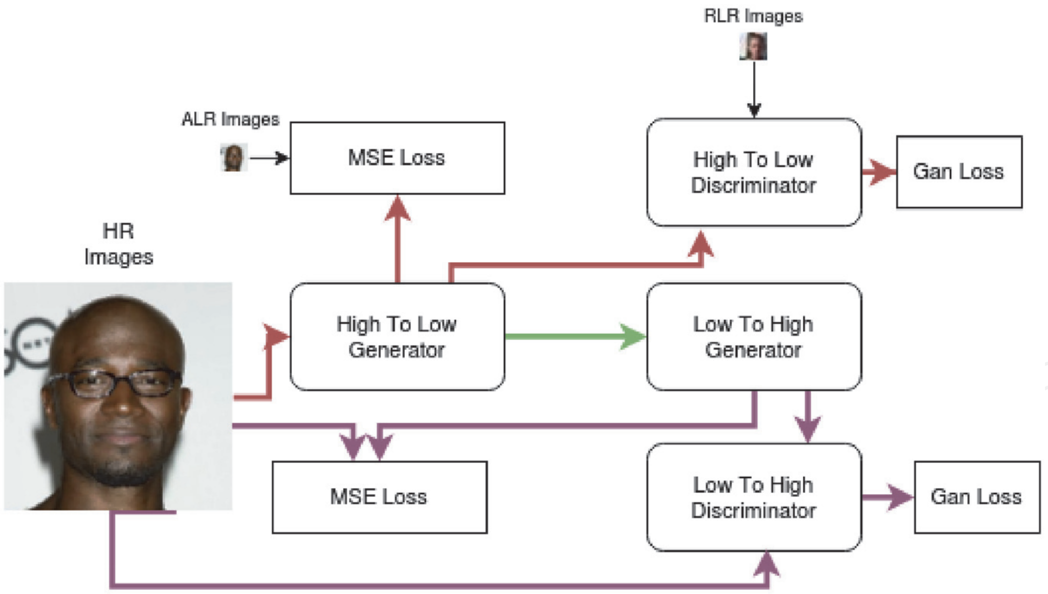


Figure 10.
Architecture of our unsupervised approach for real-world face superresolution.

Based on this observation, we do not need to enforce the cycle consistency for the direction of low-to-high transfer. Our overall network architecture is shown in **Figure 10**, consisting of two generative adversarial networks called high-to-low and low-to-high, respectively. The high-to-low GAN takes a HR face as the input and project it into the style of the real world LR faces. The low-to-high GAN takes the output of high-to-low generator as the input and try to reconstruct the original HR faces.

3.2 Dataset collection

High Resolution (HR) data: We have created our HR dataset by combining several publicly available HR face datasets: CelebA [31], AFLW [32], LS3D-W [33], and VGGFace2 [34]. For the reason of consistency, we have used S^3fd [35] to crop the face region in each image. We ended up with a total of 229,041 training images and 8892 testing images. All images are resized to 128 × 128. **Real Low Resolution Data (RLR):** We created our real LR dataset from Widerface [36] and we crop the face region using [35]. We have ended up with a total of 156,557 LR training images and 8241 LR testing images. All images have been resized to 16 × 16 (i.e., a scaling factor of 8).

Artificial Low Resolution (ALR) data: To create this dataset, we downsample our HR images by a factor of 8 using the "bilinear" method provided by Matlab. Note that the use of ALR is only for supervised learning experiences which require paired HR-LR training data.

3.3 Details of network architecture

In this section, we go through the detailed of the convolution neural networks that form our unsupervised architecture in **Figure 10**. We have two generators and two discriminators.

3.3.1 Building blocks

Our generators are made up of a number of blocks that we call residual + attention [37] (**Figure 11**). We use self attention layers as defined in [38] as part of both low-to-high and high-to-low discriminators. The details of self-attention layer are shown in **Figure 12**.

3.3.2 High-to-low generator

High-to-low generator has an encoder-decoder type architecture [39]: with the encoder consisting of five residual + attention blocks each followed by average

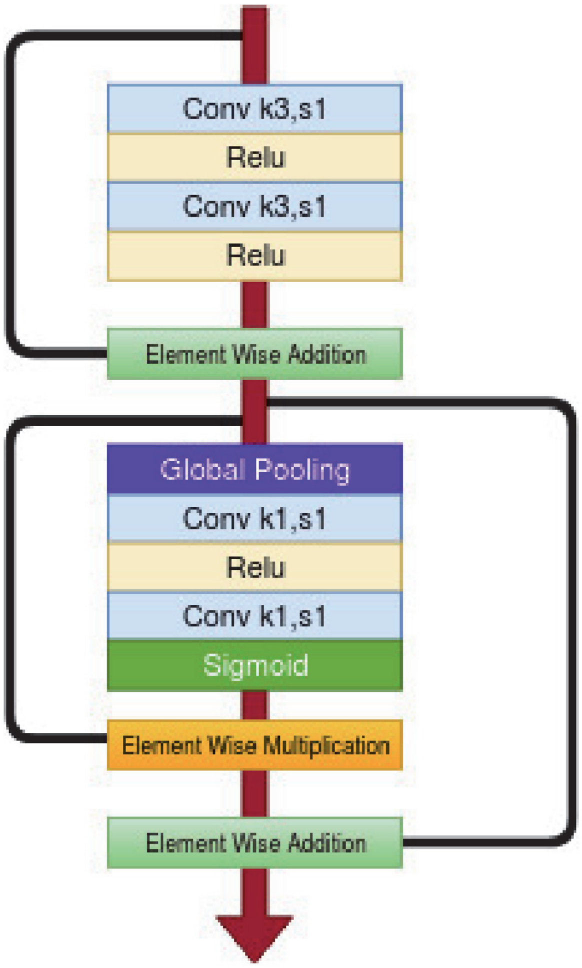


Figure 11.
Residual + attention block.

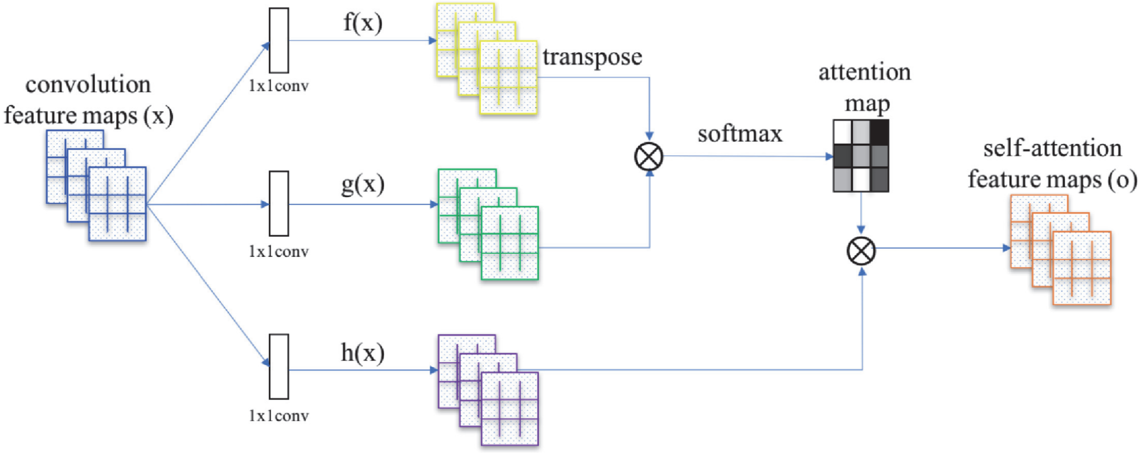


Figure 12.
Self-attention layer.

pooling layer and the decoder consisting of four residual + attention blocks where the first two are followed by bilinear upsampling layer. Therefore, the input is downsampled by a factor of 32 and upsampled by a factor of 4, which produces a down-sampled image by a factor of 8 but with more flexibility of modeling degradation (e.g., unknown blur [40]). We also concatenate a noise vector to the input image of the network using a fully connected layer, which contributes to the robustness of the proposed degradation modeling. The details of the high-to-low generator architecture is shown in **Figure 13**.

IntechOpen

3.3.3 Low-to-high generator

Low to high generator subnetwork consists of four sections of 6, 3, 2, and 1 successive residual attention blocks separated by bilinear upsampling of 2 (similar to the strategy of progressive growing GAN [25]); overall the input 16 × 16 patch is up-sampled by a factor of 8. The details of the architecture are shown in **Figure 14**.

3.3.4 High-to-low discriminator

High-to-low discriminator subnetwork consists of three convolution layers followed by a leaky relu layer and a last convolutional layer. We have added two self-attention layers at the end of the network (refer to **Figure 15**).

3.3.5 Low-to-high discriminator

Low-to-high discriminator consists of four convolution layers followed by a leaky relu layer and a last convolution layer. Similarly, we have also added two self-attention layers. The details of the architecture are in **Figure 16**.

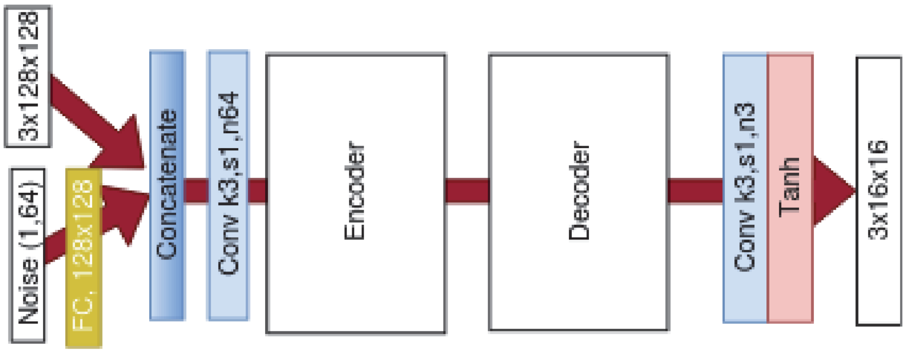


Figure 13.
Architecture of the high to low generator.

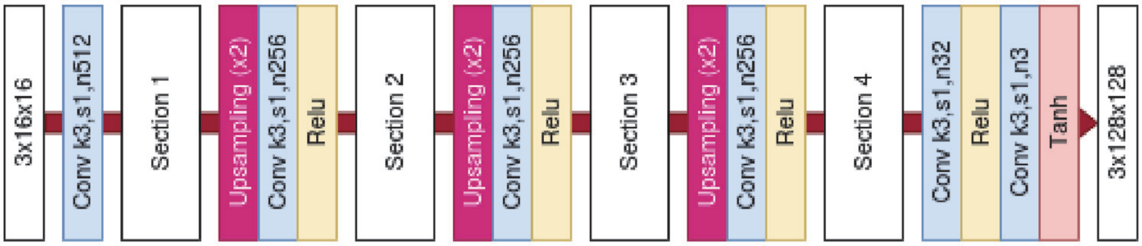


Figure 14.
Architecture of the low to high generator.

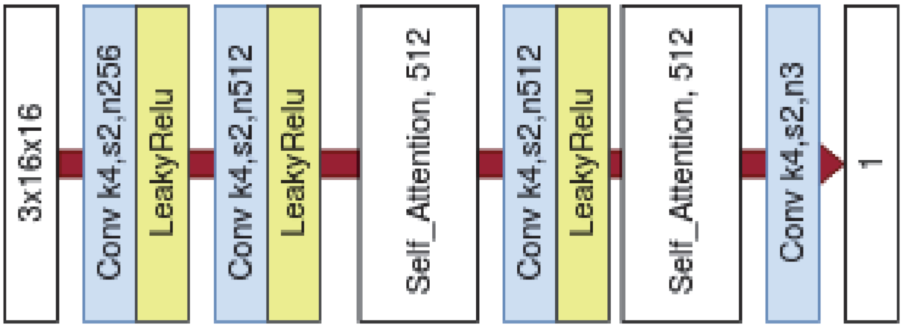


Figure 15.
Architecture of the high to low discriminator.

IntechOpen

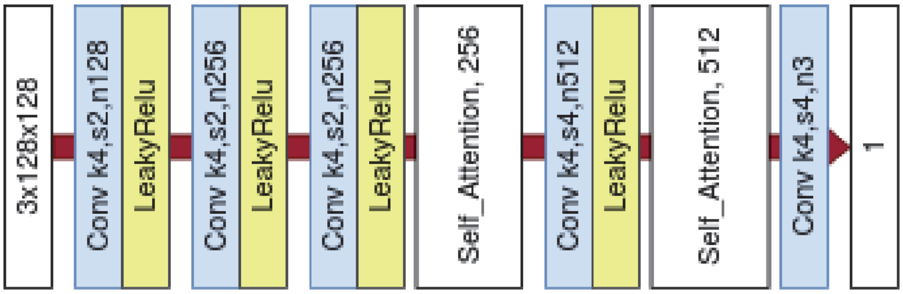


Figure 16.
Architecture of the low-to-high discriminator.

3.4 Loss functions

The generator loss, in both high-to-low and low-to-high GANs, is the weighted sum of the content loss and the GAN loss, as shown in Eq. (1) where $\alpha = 1$ and $\beta = 0.001$.

$$L_G = \alpha L_{pixel} + \beta L_{GAN}^{G} \tag{1}$$

The GAN losses and the pixel loss function follow the formula in Eqs. (2) and (3).

$$\begin{aligned} f(u,v) \quad &= \frac{1}{2}\left\{ \mathop{\mathbb{E}}_{u\sim P_u}\left[\max\left(0, 1-\left(D(u) - \mathop{\mathbb{E}}_{v\sim P_v}[D(v)]\right)\right)\right] + \right. \\ &\left. \mathop{\mathbb{E}}_{v\sim P_v}\left[\max\left(0, 1+\left(D(v) - \mathop{\mathbb{E}}_{u\sim P_u}[D(u)]\right)\right)\right]\right\} \end{aligned} \tag{2}$$

$$g(u,v) = \frac{1}{WH}\sum_{i=1}^{W}\sum_{j=1}^{H}\left(u_{i,j} - v_{i,j}\right)^2 \tag{3}$$

3.4.1 High-to-low GAN loss functions

Generator loss: As mentioned above, the generator loss is the weighted sum of the content loss and the GAN loss, Eq. (1), where $L_{GAN}^{G} = f(I_{RLR}, I_{FLR})$ and $L_{pixel} = g(I_{ALR}, I_{FLR})$.

Discriminator loss: The discriminator is defined as follows: $L_{GAN}^{D} = f(I_{FLR}, I_{RLR})$. I_{ALR} is the artificial low resolution image, I_{FLR} the fake low resolution image generated by the high to low generator, and I_{RLR} the real world low resolution images. Functions f and g are defined, respectively, in Eqs. (2) and (3).

3.4.2 Low-to-high GAN loss functions

Generator Loss: Similarly, the generator loss is the weighted sum of the content loss and the GAN loss in Eq. (1), where $L^{G}_{GAN} = f(I_{HR}, I_{FHR})$ and $L_{pixel} = g(I_{HR}, I_{FHR})$.

Discriminator loss: The discriminator is defined as follows: $L^{D}_{GAN} = f(I_{FHR}, I_{HR})$. I_{FHR} is the fake high resolution image generated by the low to high generator and I_{HR} the real world high resolution image. Functions f and g are defined, respectively, in Eqs. (2) and (3).

3.5 Training strategy

It is worth mentioning that we have not augmented the data during training by standard techniques such as image flipping, scaling, and rotation. Our experience suggests that for unsupervised learning, data augmentation does not help improve the accuracy of face SR reconstruction but increase the computational burden as well as the risk of introducing artifacts (due to unpaired LR-HR training data). We have also found that the popular normalization tricks (e.g., batch normalization [41] and spectral normalization [42]) do not help in the unsupervised scenario but have the tendency of introducing artifacts to super-resolved images.

We have used a batch of size 32, and the total training requires about 20 epochs or ~143,000 generators and discriminators updates. The learning rate is kept at 0.001 throughout the training process, and the overall architecture is trained in an end-to-end manner. We also use Adam optimizer [43] with $\beta_1 = 0$ and $\beta_2 = 0.9$ and adopt a PyTorch-based implementation [44].

4. Experimental results

In this section, we present a comparison between our style-based approaches toward SISR of face images and state-of-the-art supervised (FSRNET [21]) and unsupervised ones (Bulat's [3]). First, we use extensive ablation studies to show the effect of removing the low-to-high discriminator (**Figure 16**) from the architecture and demonstrate the output of our degradation model. As an extension of our ablation study, we also present a supervised approach for ALR face images based on using GAN composed of our low-to-high generator (**Figure 14**) and low-to-high discriminator (**Figure 16**) but without any cycle involved. Finally, we will report our unsupervised learning results on real-world LR face images and compare them against other competing approaches.

4.1 Ablation study

4.1.1 Importance of the discriminator

We have compared the outputs of our unsupervised architecture with and without the high-to-low discriminator. The image comparison results are shown in **Figure 17**. It is obvious that discriminator plays a significant role in improving the visual quality of super-resolved images.

4.1.2 Degradation modeling

We next show the capability of high-to-low network on learning real-world degradation models. **Figure 18** includes several typical examples of learned LR images from HR inputs. It is important to note that our high-to-low network has managed to learn a variety of degradation models including varyimg poses and severe blurs.

4.1.3 Deep features visualization

We also visualize the feature maps of the low-to-high generator in **Figure 14**. In this visualization experiment, we have plotted the output from Sections 2, 3, and 4 as shown in **Figure 19**. It can be observed that as section/layer number increases, the learned feature representations have a larger field of view as well as more sophisticated semantic information related to faces.

4.2 Supervised approach

We have experimented with a GAN-based supervised approach toward SISR as an extension of our ablation study. Such experiment is included to demonstrate a degeneration of network architecture from unsupervised (**Figure 10**) to supervised (**Figure 20**) setting. We have used the same architecture for the low-to-high

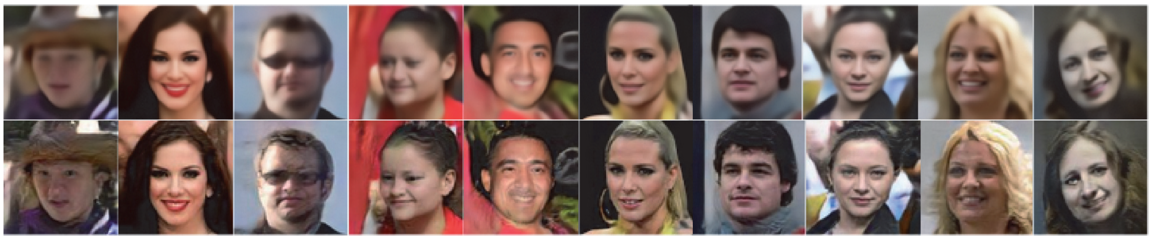


Figure 17.
Comparison between our unsupervised method with and without the discriminator on Widerface dataset [36]. First row without discriminator; second row with discriminator. Please zoom in for better visualization.

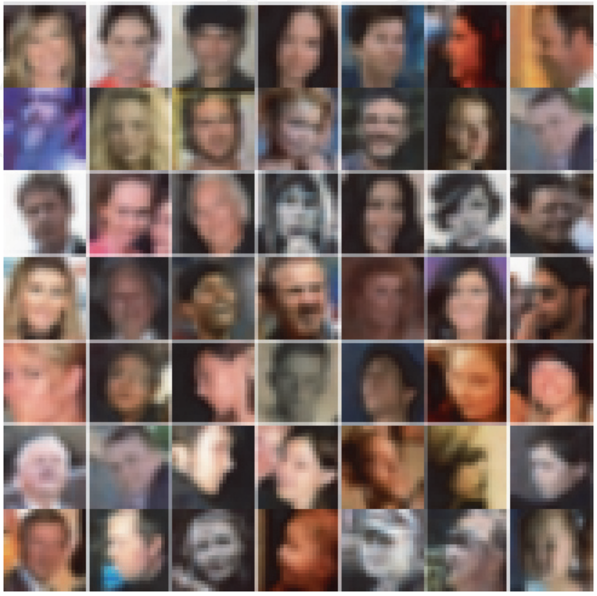


Figure 18.
Effectiveness of degradation model learning: exemplar synthetic LR images from the high-to-low network (note the rich variability and similarity to the real-world Widerface dataset [36]).

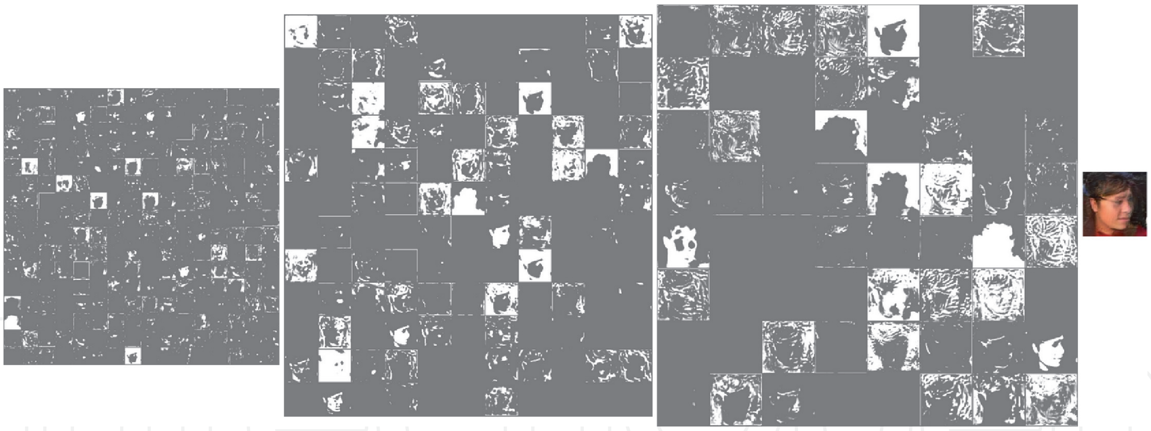


Figure 19.
Feature maps for low-to-high generator. From left to right: ×2-upsampling, ×4-upsampling, ×8-upsampling, and the output image.

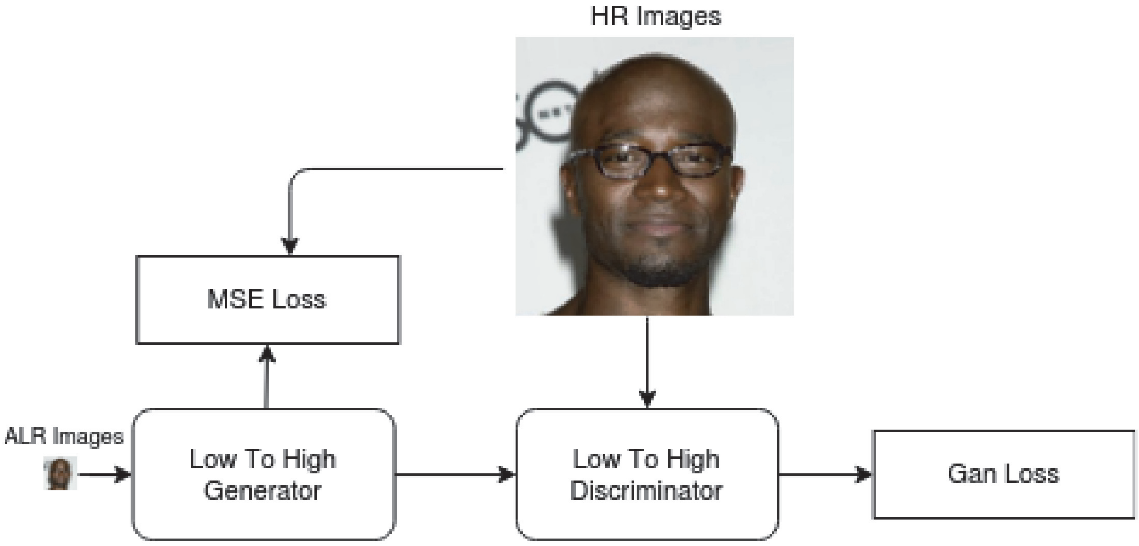


Figure 20.
Architecture of the supervised approach.

generator and discriminator as in **Figures 14** and **16**. To get paired LR and HR images; we downsample the original high resolution faces by a scaling factor of 8 to create artificial low resolution (ALR) images as explained in Section 3.2. The overall architecture of this reduced supervised approach is detailed in **Figure 20**.

4.2.1 Loss functions

Similar to Section 3.4, the loss functions for our supervised approach are defined as follows:

Generator loss: the generator loss is the weighted sum of the content loss and the GAN loss, Eq. (1) where $L_{GAN}^{G} = f(I_{FHR}, I_{HR})$ and $L_{pixel} = g(I_{HR}, I_{FHR})$.

Discriminator loss: $L_{GAN}^{D} = f(I_{HR}, I_{FHR})$ where I_{HR} denotes the high resolution image and I_{FHR} is the reconstructed high resolution one generated by our network. Functions f and g are defined, respectively, in Eqs. (2) and (3).

4.2.2 Training strategy

We have used a batch of size 32 and trained for 20 epochs or ~143,000 updates of generator and discriminator. The learning rate is kept at $1e - 4$ throughout the training process. We have used Adam optimizer [43] with $\beta_1 = 0$ and $\beta_2 = 0.9$ and implemented our supervised learning SR using Pytorch [44].

4.3 Comparison with other supervised approaches

We compare our unsupervised face super-resolution approach with two supervised approaches; FSRNET [21] and our own approach outlined in Section 4.2. FSRNET [21] uses geometric priors to estimate (e.g., facial landmark heat maps and parsing maps) to facilitate the procedure of supervised learning.

4.3.1 Performance on artificial low resolution test data

We report our experimental results for ALR data and compare them against the current state-of-the-art FSRNet/FSRGAN [22]. Despite being synthetic, ALR images are still useful because they have ground-truth (HR) available and appropriate for gauging the performance of supervised learning (with paired HR-LR training data).

Subjective quality comparisons: Figure 21 shows the qualitative comparisons between our supervised/unsupervised approaches and state-of-the-art supervised method FSRGAN [21]. It can be easily verified that ours can produce visually more convincing and pleasant HR results than FSRGAN (e.g., sharper contrast, more natural hair, and fewer artifacts around earrings).

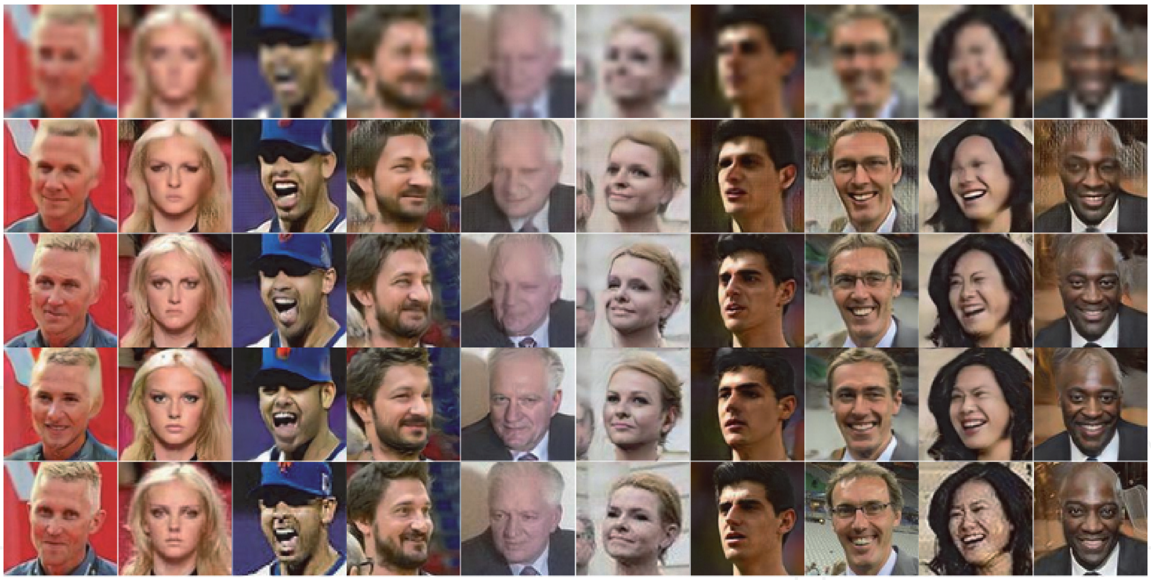


Figure 21.
Visual quality comparisons among competing methods on artificial low resolution face images. Rows top-down: bicubic, FSRGAN [21], ours (supervised), ours (unsupervised) and groundtruth. Please zoom in for better visualization.

Method	PSNR
FSRGAN [21]	22.840
Ours (Supervised)	**23.65**
Ours (Unsupervised)	21.97

Table 1.
Objective quality results of different methods on ALR images in terms of PSNR (dB) (highest PSNR is highlighted by bold-face).

Objective quality comparisons: We report the comparison results in terms of peak signal-to-noise ratio (PSNR) in **Table 1**. Our supervised learning outperforms FSRGAN [21] by as much as 0.8*dB*.

4.3.2 Performance on real world wider test data

We tested both our supervised and unsupervised methods on the popular real-world LR dataset Widerface [36]. This dataset is particularly challenging for face detection and SR because its 393,703 faces contain a high degree of variability in scale, pose, and occlusion. Due to lack of groundtruth images (HR counterparts) for this dataset, we have to count on visual quality comparison alone for performance evaluation (without PSNR comparisons). We report our visual quality comparison between our methods and current state-of-the-art supervised (FSRGAN [22]). **Figure 22** shows the results; we can see that the unsupervised approach outperforms supervised ones in case of real world low resolution images.

4.4 Comparison with state-of-the-art unsupervised approach

We compare our unsupervised method with Bulat's [45]. We show that our methods are able to better preserve facial features **Figure 23**.

One can observe that the SR image produced by Bulat's [45] method has the following problems: age variation (third), gender swapping (fourth and seventh), and artifacts (second and seventh).

4.5 Performance in term of receiver operating curve (ROC)

Using Openface [46] matching algorithm, we plot the ROC curve for three types of degradation models: artificial degradation (or ALR), jpeg compression, and our high-to-low degradation model. We show that our proposed supervised approach outperform all previous ones in case of artificial degradation and our unsupervised approach performs the best when it comes to the other two degradation models.

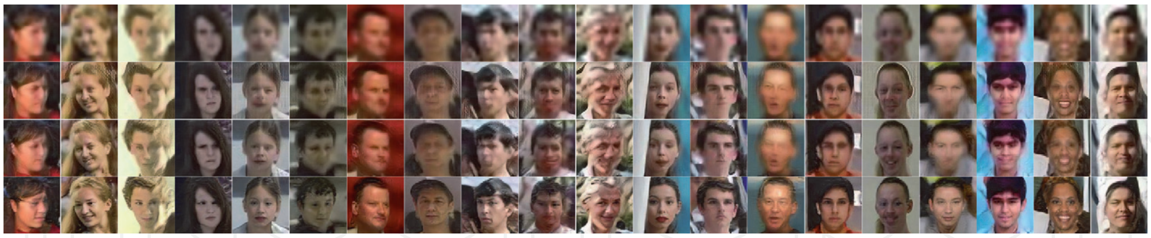


Figure 22.
Qualitative comparisons with FSRGAN on Widerface test data. Rows are respectively: Bicubic, FSRGAN, ours (supervised), and ours (unsupervised). Please zoom in for better visualization.

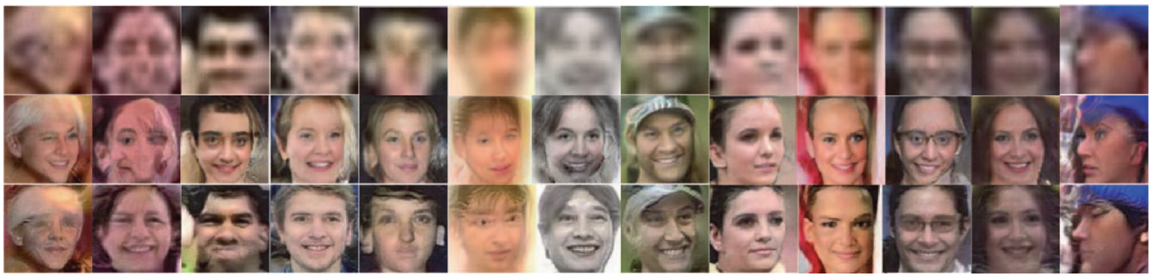


Figure 23.
Qualitative comparisons with Bulat's method [3] on Widerface test data. Rows are respectively: Bicubic, Bulat's method, and ours (unsupervised). Please zoom in for better visualization.

4.5.1 Performance on artificial low resolution data

We have compared our supervised and unsupervised approach with FSRNET [21] and Bulat's [45] in terms of ROC curve results. We use our Artificial Low Resolution (ALR) test data. **Figure 24** shows that our supervised method outperforms all other methods. On the other hand, our unsupervised method performs worse than FSRNET [21] but still performs significantly better than previous unsupervised state-of-the-art approach Bulat's [45].

IntechOpen

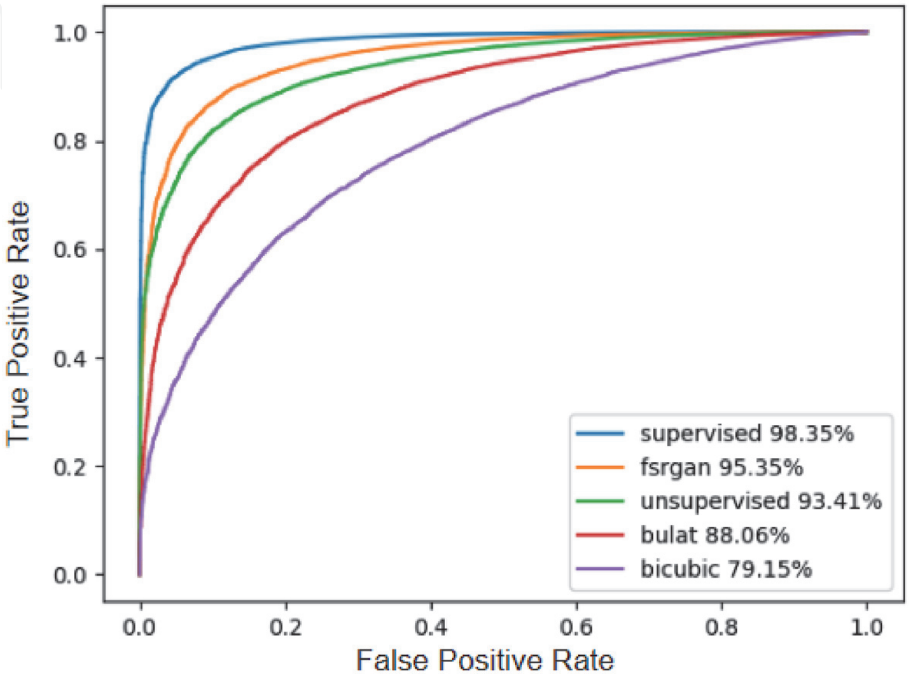


Figure 24.
ROC curve for ALR test data.

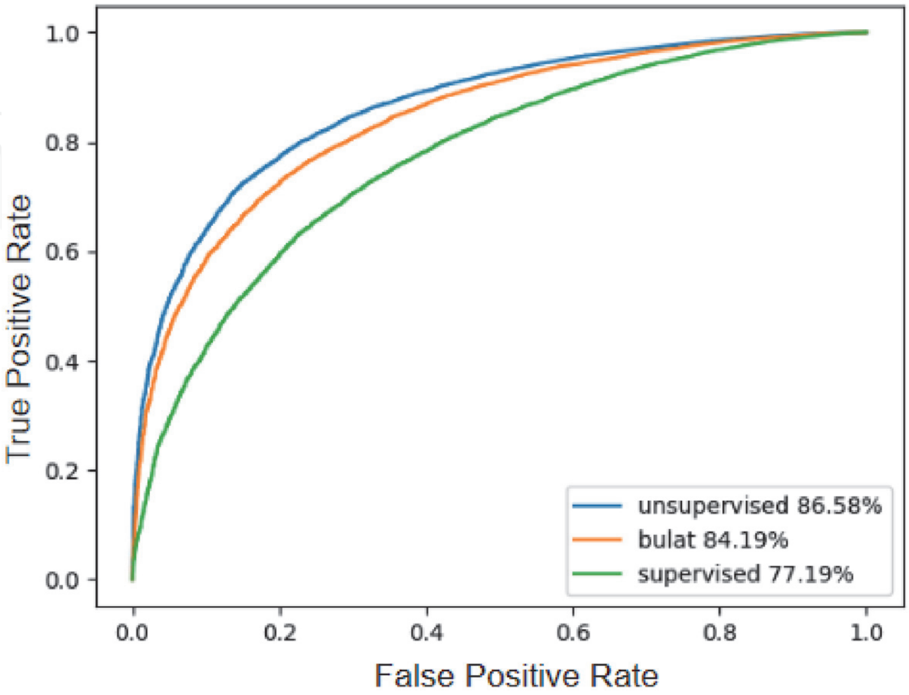


Figure 25.
ROC curve for compressed test data.

4.5.2 Performance on compressed data

We compressed our Artificial Low Resolution (ALR) test data using JPEG lossy compression. We used this compressed data to plot the ROC curve for our supervised and unsupervised approach as well as Bulat's [3]. Our unsupervised approach outperforms better than the other ones, as shown in **Figure 25**.

4.5.3 Performance on generated low resolution data

We also plotted the ROC curve using the data generated by passing high resolution test data to our trained high-to-low generator. We show here that our unsupervised approach performs better than our supervised one; **Figure 26**.

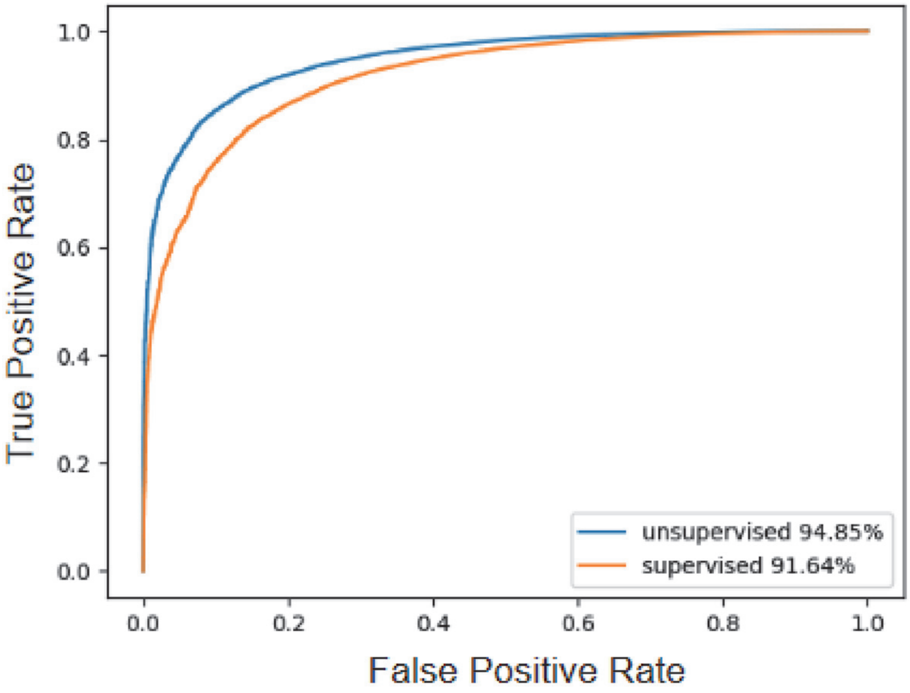


Figure 26.
ROC curve for our high-to-low degradation model.

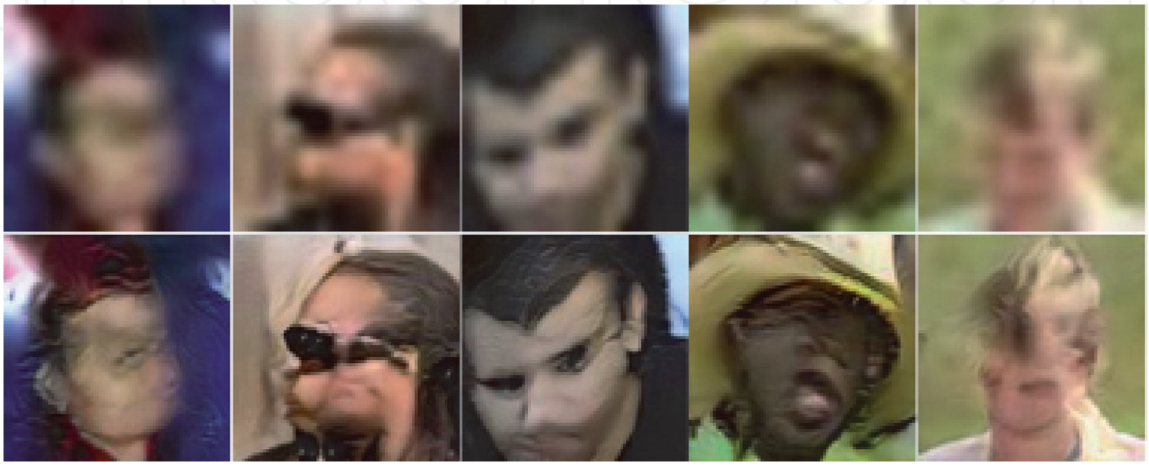


Figure 27.
Exemplar failure cases of our unsupervised method on Widerface test data. Top-row: Input LR; bottom-row: our SR result (unsupervised).

4.6 Failure cases

As mentioned above, our approach intentionally skips the step of data augmentation. It turns out that our method is still sensitive to extreme variations of face pose such as as those shown in **Figure 27**. Due to severe occlusions and large pose variations, those LR examples are often rare even among Widerface dataset. This is within our expectation because high-to-low network simply does not have sufficient training data to learn the challenging degradation model. Note that similar findings have been reported for Bulat's method in [3] (refer to **Figure 9** in that paper).

5. Conclusions

We have studied the problem of SISR for real-world face images in this chapter and presented an unsupervised learning approach toward such blind reconstruction of SR images. The challenging scenario of real-world LR low-quality images defies conventional approaches based on paired HR-LR training data because groundtruth HR images are generally unavailable for real-world LR images. By pairing style-based generator with relativistic discriminator, we demonstrate an unsupervised learning approach with GAN-based end-to-end optimization that is capable of advancing the state-of-the-art in blind SR reconstruction of real-world LR face images. We have compared our degradation modeling against previous Bulat's method as well as their ROC performance on artificial LR dataset. Extensive experimental results have shown favorable performance for the proposed method over Bulat's method.

Author details

Ahmed Cheikh Sidiya and Xin Li*
Lane Department of Computer Science and Electrical Engineering, Morgantown, USA

*Address all correspondence to: xin.li@ieee.org

IntechOpen

© 2020 The Author(s). Licensee IntechOpen. Distributed under the terms of the Creative Commons Attribution - NonCommercial 4.0 License (https://creativecommons.org/licenses/by-nc/4.0/), which permits use, distribution and reproduction for non-commercial purposes, provided the original is properly cited. CC BY-NC

References

[1] Goodfellow I, Bengio Y, Courville A. Deep Learning. Cambridge, Massachusetts: MIT Press; 2016

[2] Christian L, Lucas T, Ferenc H, Jose C, Andrew C, Alejandro A, 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. pp. 4681-4690

[3] Adrian B, Jing Y, Georgios T. To learn image super-resolution, use a gan to learn how to do image degradation first. In: Proceedings of the European Conference on Computer Vision (ECCV); 2018. pp. 185-200

[4] Zhu J-Y, Park T, Isola P, Efros AA. Unpaired image-to-image translation using cycle-consistent adversarial networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2017. pp. 2223-2232

[5] Karras T, Laine S, Aila T. A style-based generator architecture for generative adversarial networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2019. pp. 4401-4410

[6] Jolicoeur-Martineau A. The relativistic discriminator: a key element missing from standard gan. In: International Conference on Learning Representations; 2019

[7] Dong C, Loy CC, He K, Tang X. Image super-resolution using deep convolutional networks. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2015;**38**(2):295-307

[8] Fukushima K, Miyake S. Neocognitron: A self-organizing neural network model for a mechanism of visual pattern recognition. In: Competition and Cooperation in Neural Nets. Springer; 1982. pp. 267-285

[9] Vinod N, Geoffrey EH. Rectified linear units improve restricted boltzmann machines. In: Proceedings of the 27th International Conference on Machine Learning (ICML-10); 2010. pp. 807-814

[10] Dominik S, Andreas M, Sven B. Evaluation of pooling operations in convolutional architectures for object recognition. In: International Conference on Artificial Neural Networks. Springer; 2010. pp. 92-101

[11] Huang TS, Yang J, Wright J, Ma Y. Image super-resolution via sparse representation. IEEE Transactions on Image Processing. 2010;**19**(11): 2861-2873. abs/1501.00092

[12] He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016. pp. 770-778

[13] Chao D, Chen CL, Xiaoou T. Accelerating the super-resolution convolutional neural network. In: European Conference on Computer Vision. Springer; 2016. pp. 391-407

[14] Kim J, Lee JK, Lee KM. Deeply-recursive convolutional network for image super-resolution. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016. pp. 1637-1645

[15] Kim J, Lee JK, Lee KM. 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

[16] Bee L, Sanghyun S, Heewon K, Seungjun N, Kyoung ML. Enhanced deep residual networks for single image super-resolution. In: Proceedings of the

IEEE Conference on Computer Vision and Pattern Recognition Workshops. 2017. pp. 136-144

[17] Wei-Sheng L, Jia-Bin H, Narendra A, Ming-Hsuan Y. Deep laplacian pyramid networks for fast and accurate super-resolution. In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2017

[18] Yulun Z, Yapeng T, Yu K, Bineng Z, Yun F. Residual dense network for image super-resolution. CVPR; 2018

[19] Yulun Z, Kunpeng L, Kai L, Lichen W, Bineng Z, Yun F. Image super-resolution using very deep residual channel attention networks. ECCV; 2018

[20] Adrian B, Georgios T. Super-fan: Integrated facial landmark localization and super-resolution of real-world low resolution faces in arbitrary poses with gans. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018. pp. 109-117

[21] Yu C, Ying T, Xiaoming L, Chunhua S, Jian Y. 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

[22] Goodfellow IJ, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, et al. Generative adversarial nets. In: Proceedings of the 27th International Conference on Neural Information Processing Systems - Volume 2, NIPS'14. Cambridge, MA, USA: MIT Press; 2014. pp. 2672-2680

[23] Alec R, Luke M, Soumith C. Unsupervised representation learning with deep convolutional generative adversarial networks. In: International Conference on Learning Representations (ICLR); 2015

[24] Han Z, Ian G, Dimitris M, Augustus O. Self-attention generative adversarial networks. arXiv preprint arXiv: 1805.08318; 2018

[25] Tero K, Timo A, Samuli L, Jaakko L. Progressive growing of gans for improved quality, stability, and variation. CoRR. abs/1710.10196; 2017

[26] Tero K, Samuli L, Miika A, Janne H, Jaakko L, Timo A. Analyzing and improving the image quality of stylegan. arXiv preprint arXiv:1912.04958; 2019

[27] Miles B, Shahar A, Jack C, Helen T, Peter E, Ben G, et al. The malicious use of artificial intelligence: Forecasting, prevention, and mitigation. ICLR; 2018

[28] Phillip I, Jun-Yan Z, Tinghui Z, Alexei AE. Image-to-image translation with conditional adversarial networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2017. pp. 1125-1134

[29] Jun-Yan Z, Taesung P, Phillip I, Alexei AE. Unpaired image-to-image translation using cycle-consistent adversarial networks. In: Proceedings of the IEEE International Conference on Computer Vision. 2017. pp. 2223-2232

[30] Mehdi M, Simon O. Conditional generative adversarial nets. arXiv preprint arXiv:1411.1784; 2014

[31] Ziwei L, Ping L, Xiaogang W, Xiaoou T. Deep learning face attributes in the wild. In: Proceedings of International Conference on Computer Vision (ICCV); December 2015

[32] Roth Martin Koestinger PM, Paul W, Horst B. Annotated facial landmarks in the wild: A large-scale, real-world database for facial landmark localization. In: Proceedings of the First IEEE International Workshop on Benchmarking Facial Image Analysis Technologies; 2011

[33] Adrian B, Georgios T. How far are we from solving the 2d & 3d face

alignment problem? (and a dataset of 230,000 3d facial landmarks). In: International Conference on Computer Vision; 2017

[34] Cao Q, Shen L, Xie W, Parkhi OM, Zisserman A. Vggface2: A dataset for recognising faces across pose and age. In: International Conference on Automatic Face and Gesture Recognition. 2018

[35] Shifeng Z, Xiangyu Z, Zhen L, Hailin S, Xiaobo W, Stan ZL. S3 fd: Single shot scale-invariant face detector. In: Proceedings of the IEEE International Conference on Computer Vision. pp. 192-201

[36] Shuo Y, Ping L, Chen CL, Xiaoou T. Wider face: A face detection benchmark. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2016

[37] Yulun Z, Kunpeng L, Kai L, Lichen W, Bineng Z, Yun F. Image super-resolution using very deep residual channel attention networks. In: Proceedings of the European Conference on Computer Vision (ECCV). pp. 286-301

[38] Zhang H, Goodfellow I, Metaxas D, Odena A. Self-attention generative adversarial networks. In: International Conference on Machine Learning. 2019. pp. 7354-7363

[39] Vijay B, Alex K, Roberto C. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. arXiv preprint arXiv:1511.00561 5; 2015

[40] Jinjin G, Hannan L, Wangmeng Z, Chao D. Blind super-resolution with iterative kernel correction. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2019. pp. 1604-1613

[41] Sergey I, Christian S. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In: International Conference on Machine Learning. 2015. pp. 448-456

[42] Takeru M, Toshiki K, Masanori K, Yuichi Y. Spectral normalization for generative adversarial networks. ICLR; 2018

[43] Diederik PK, Jimmy B. Adam: A method for stochastic optimization, 2014. cite arxiv:1412.6980. Comment: Published as a conference paper at the 3rd International Conference for Learning Representations. San Diego; 2015

[44] Adam P, Sam G, Francisco M, Adam L, James B, Gregory C, et al. Pytorch: An imperative style, high-performance deep learning library. In: Wallach H, Larochelle H, Beygelzimer A, dAlché-Buc F, Fox E, Garnett R, editors. Advances in Neural Information Processing Systems. Curran Associates, Inc.; 2019. pp. 8024-8035

[45] Adrian B, Jing Y, Georgios T. To learn image super-resolution, use a gan to learn how to do image degradation first. ECCV; 2018

[46] Brandon A, Bartosz L, Mahadev S. Openface: A general-purpose face recognition library with mobile applications. Technical report, CMU-CS-16-118, CMU School of Computer Science; 2016