

# PCA-SRGAN: Incremental Orthogonal Projection Discrimination for Face Super-resolution

Hao Dou<sup>1,2</sup>, Chen Chen<sup>1,2</sup>, Xiyuan Hu<sup>3\*</sup>, Zuxing Xuan<sup>4</sup>, Zhisen Hu<sup>5</sup>, Silong Peng<sup>1,2</sup>

<sup>1</sup>Institute Of Automation, Chinese Academy Of Sciences

<sup>2</sup>The Chinese academy of science

<sup>3</sup>Nanjing University of Science and Technology

<sup>4</sup>Beijing Union University

<sup>5</sup>Beijing University of Posts and Telecommunications

{douhao2015, chen.chen,silong.peng}@ia.ac.cn, xiyuan.hu@foxmail.com, zuxingxuan@163.com, huzhisen1117@gmail.com

## Abstract

Generative Adversarial Networks (GAN) have been employed for face super resolution but they bring distorted facial details easily and still have weakness on recovering realistic texture. To further improve the performance of GAN-based models on super-resolving face images, we propose PCA-SRGAN which pays attention to the cumulative discrimination in the orthogonal projection space spanned by PCA projection matrix of face data. By feeding the principal component projections ranging from structure to details into the discriminator, the discrimination difficulty will be greatly alleviated and the generator can be enhanced to reconstruct clearer contour and finer texture, helpful to achieve the high perception and low distortion eventually. This incremental orthogonal projection discrimination has ensured a precise optimization procedure from coarse to fine and avoids the dependence on the perceptual regularization. We conduct experiments on CelebA and FFHQ face datasets. The qualitative visual effect and quantitative evaluation have demonstrated the overwhelming performance of our model over related works.

## 1 Introduction

Face image super-resolution (FSR), also known as face hallucination, aims to recover a high-resolution (HR) face image from its low-resolution (LR) input, which is helpful to numerous facial analysis applications. For many years image super-resolution (SR) has been a challenging research problem and data-driven deep learning draws a lot of attention recently. Deep learning has been employed to achieve an end-to-end single image super-resolution model, such as the Convolutional Neural Networks (CNN) based methods (e.g. SRCNN [Dong *et al.*, 2014], SRResnet [Kim and Lee, 2018]) and recent approaches using Generative Adversarial Networks (GAN) [Goodfellow *et al.*, 2014] (e.g. SRGAN [Ledig

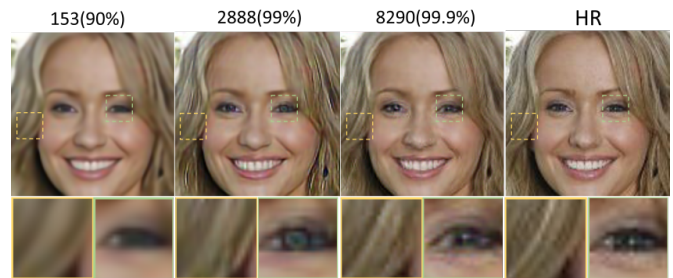


Figure 1: The reconstruction results of our PCA-SRGAN by incremental discrimination along with the gradually increased subspace. Above the images list the corresponding subspace dimensions and energy proportions.

*et al.*, 2017], ESRGAN [Wang *et al.*, 2018a]), which outperform most of traditional methods. Specifically for face image super resolution task, many types of face priors are combined and applied to deep-learning-based SR methods. Most of them learn a straight mapping by a variety of regularizations, including facial masks [Chen *et al.*, 2018], attributes [Yu *et al.*, 2018], etc.

Compared to CNN-based models which aim at the less distortion, GAN-based SR models have demonstrated the superior perceptual performance. Extra perceptual loss like VGG [Johnson *et al.*, 2016; Ledig *et al.*, 2017] are often accompanied to regularize the generator for more stable results with less artifacts. However, GAN-based models still have weakness on generating realistic and precise texture and produce fake details easily especially for face images which lead to the distortion. To enhance the performance of GAN-based models, cumulative learning is considered as a effective strategy and introduced to super resolution task to lower the reconstruction difficulty. Typical cumulative learning for super resolution concentrates on the incrementally network structure – progressively grows both the networks and the up-scale factors [Wang *et al.*, 2018b] or pyramid levels [Lai *et al.*, 2017]. Different from this kind of cumulative training in which the whole networks grow for different stages, we can maintain the network structure and seek a new cumulative

\*Corresponding Author

training method for adversarial learning — incrementally discriminate the different levels of information or components for easier and more precise reconstruction.

As one of the classical statistical methods, principal component analysis (PCA) [Jolliffe, 1986] can give a perfect hierarchical representation of face images, namely the orthogonal projections on the PCA orthogonal subspace. We can train the GAN-based SR model by discriminating the projections of face images on the PCA subspace cumulatively to reconstruct the super-resolved images. Therefore we propose a novel cumulative discrimination and reconstruction approach in this paper referred as *PCA-SRGAN* and focus on the *Incremental Orthogonal Projection Discrimination* in the PCA orthogonal subspace. The orthogonal projections are obtained by projecting both generated SR images and HR images into the PCA subspace. The corresponding orthogonal projections containing hierarchical facial components from structure to details will be fed into the discriminator to guide the precise optimization and fine-grained reconstruction of generator. In this way, our proposed model can lighten the difficulty of discrimination, stabilize the procedure of adversarial training and enhance the performance of generator, which ensures the ability to produce perceptual and realistic texture while brings relatively low level of distortion and artifacts without any help of auxiliary perceptual regularization. An example of cumulative training process has been shown in Figure 1. With the growth of subspace proportion, our model has generated the super-resolved face with more realistic details gradually.

In summary, the main contributions of this work are presented as follows:

- we propose a novel cumulative learning strategy for the GAN-based SR method utilizing the incremental orthogonal projection discrimination in the PCA subspace to enhance the face SR task.
- Our model provides a precise and stable adversarial training method without the need for auxiliary perceptual (VGG) loss, which achieves compelling visual effect with better perception-distortion trade-off than other methods.

## 2 Related Works

### 2.1 Face super resolution

Face super-resolution, is a special case of single image super-resolution (SISR). The SISR problem has been extensively studied in the literature using a variety of deep-learning-based models. SRCNN, EDSR [Lim *et al.*, 2017] and SRResnet employ deep convolutional network with various structures for super-resolution. Many works, e.g. SRGAN, ES-RGAN, EPSR [Vasu *et al.*, 2018], and RankSRGAN [Zhang *et al.*, 2019], also introduce the generative adversarial network to produce perception-oriented results. The challenge PIRM2018 [Blau *et al.*, 2018] has been also held to evaluate the SR methods by jointly quantifying the accuracy and perceptual quality, namely the trade-off between distortion and perception. Based on commonly used SISR methods above, the facial priors are often exploited and combined for face super resolution. There are many types of prior such as

masks, landmarks, heatmaps, applied to the face super resolution [Yang *et al.*, 2018; Song *et al.*, 2017]. Yang decomposes face images into facial components which can be super-resolved separately and fused to the complete face. Adrian Bulat learns to localize the facial landmarks as a constraint for face super resolution [Bulat and Tzimiropoulos, 2018]. Yu Chen employs an extra network to predict the prior such like facial masks to enhance the performance for very low resolution face [Chen *et al.*, 2018]. Besides the structural prior mentioned above, statistical information has been explored for face hallucination. Principal components analysis is a classical statistical prior especially useful for human face problem. Wang replaces the base vectors of PCA dictionary for LR images with that for HR images and learn the coefficients mapping between LR and HR [Wang and Tang, 2005]. In the view of multi-scale space-frequency domain, Huang designs a wavelet-based network to reconstruct the facial high-frequency information [Huang *et al.*, 2017]. Among all kinds of face prior, PCA contains all the necessary information to build a real face and can be applied on end-to-end networks easily.

### 2.2 Cumulative learning

Cumulative learning is a effective learning strategy for complex learning task which refers to starting from an easier sub-task and gradually increasing the task difficulty. It is very suitable for the image generation task, including the image super-resolution. ProGAN [Karras *et al.*, 2017] enlarges the network and simultaneously increases the resolution of generated images for easier convergence and better image generation for GAN models. Similarly, Wang performs SR starting from  $2\times$  upsampling and then blending with the portions of  $4\times$  or larger scaling factors [Wang *et al.*, 2018b], which are progressive both on architectures and training procedure. Deokyun Kim [Kim *et al.*, 2019] introduces attention mechanism to the progressive face SR for very small face images. Different from the progressive method by increasing the upsampling scale, LapGAN [Denton *et al.*, 2015] and LapSRN [Lai *et al.*, 2017] decompose the generation process by the means of the Laplacian pyramid and train the generator progressively at every level of the pyramid. Compared to common training procedure, the curriculum learning strategy can greatly reduce the training difficulty. Inspired by this, we propose a new type of cumulative learning – incrementally learn the discrimination for the PCA projection of face images.

## 3 Proposed Method

GAN-based models have weak performance on super-resolving realistic and perceptual face images. To address the issue, we lower the learning difficulty of GAN by incrementally discriminating the SR image’s projection on the subspace spanned by the elements of the PCA dictionary generated for HR images. In this section, we firstly introduce the properties of orthogonal projection on principal component space, then describe the model structure and loss functions. At last, the overall algorithm will be summarized.

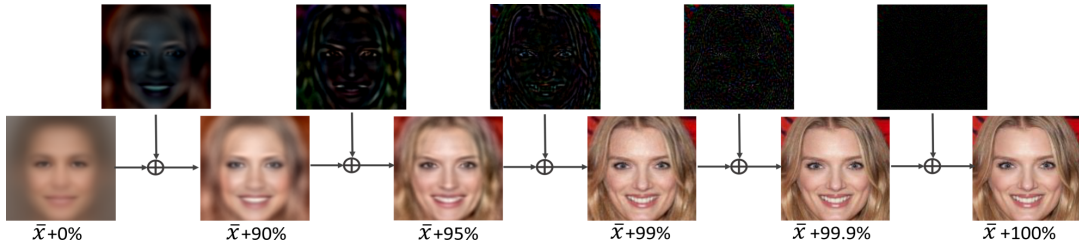


Figure 2: Visualization for different proportions of principal component projections which have been added to the average face  $\bar{x}$ . With the growth of proportion, the projections will contain increasing facial information from coarse to fine.

### 3.1 PCA orthogonal projection

Principal component analysis (PCA) is a statistical procedure using an orthogonal transformation to convert a set of observations of possibly correlated variables into a set of values of linearly uncorrelated variables called principal components.

Given a set of HR face samples  $\{x_1^h, x_2^h, x_3^h, \dots, x_m^h\}$ , among them  $x_i^h$  is a image vector which contain  $n$  pixels and  $m$  is the number of samples. Firstly,

all samples are centralized by zero-mean normalization.

$$X = [x_1^h - \bar{x}, x_2^h - \bar{x}, \dots, x_m^h - \bar{x}], \quad (1)$$

where  $\bar{x}$  denotes the average face. We construct the covariance matrix and perform the eigen decomposition.

$$X X^T P = \lambda P. \quad (2)$$

The projection matrix can be obtained from the eigenvectors of the covariance matrix.

$$P = (p_1, p_2, \dots, p_d), \quad (3)$$

where  $p_i$  is the  $i$ th eigen vector and we use  $\lambda_i$  to denote the corresponding eigen value.

A sample can be projected into the orthogonal PCA subspace as the orthogonal projection. The dimensions of orthogonal subspace are determined by the vectors' number of the projection matrix. As shown in Figure 2, the principal components, or described as the orthogonal projections of a face image on different dimensions of orthogonal subspace, has included different levels of facial information. It will be reasonable to feed these projections which represent face from structure to details into the discriminator to improve the reconstruction of super-resolved faces.

### 3.2 Incremental orthogonal projection discrimination

The pipeline is illustrated in Figure 3. We perform PCA on HR datasets and get the orthogonal space spanned by the PCA projection matrix  $P$ . The orthogonal space can be split into two orthogonal complementary subspace  $W$  and  $V$  respectively determined by  $P_W$  and  $P_V$ .  $P_W$  is a projection matrix which contains the first  $n$  vectors of  $P$  and  $P_V$  contains the rest vectors.  $G$  and  $D$  denote the generator and discriminator separately.

Given a generated SR image  $x^s$  which has been normalized by subtract the mean,  $x_W^s = P_W x^s$  is the orthogonal projection of  $x^s$  on the subspace  $W$ ,  $x_V^s = P_V x^s$  is the orthogonal projection of  $x^s$  on the subspace  $V$ . Similar operations are also applied to the HR groundtruth  $x^h$ .  $x_W^s$  will be sent to the discriminator to guide the adversarial learning for accurate face SR reconstruction. Both of the  $x_W^s$  and

$x_V^s$  will be constrained by the L1 loss. The LR input will be bicubic-upsampled as the condition to assist the discrimination of GAN. We also illustrate the incremental training procedure by the progress bar shown in the lower part of Figure 3. The vernier will slide from the left to the right side during the training procedure, which means the subspace  $W$  will expand from zero to the almost full space, while the opposite situation for subspace  $V$ . Corresponding to the visualized projections in Figure 2, the growing subspace  $W$  will absorb more and more detailed and precise face components prepared for discrimination, helpful to reduce the learning difficulty of discriminator and make the generator catch the distribution of original HR data progressively. Meanwhile, the shrinking subspace  $V$  is regarded as a repository which stores the constrained residual information to be added into  $W$ . This curriculum learning strategy guarantees a stable and precise learning process and has the ability to synthesize compelling facial details without visible distortion. To notice that, the subspace  $W$  is increased to 99.9% of the full space and the last 0.1% percentage could be treated as the random noise space and not necessary to be regularized by the GAN model.

**Networks structure** The proposed model consists of only two networks, the generator and the discriminator, which construct the elementary structure of GAN. The same to the baseline ESRGAN, RRDBnet is used as the generator which combines multilevel residual network and dense connections without BN layers, and the discriminator contains eight convolutional layers and two linear layers to generate the classification value.

### 3.3 Loss functions

We use pixel-wised L1 loss and GAN loss without any auxiliary feature-based or perceptual loss to obtain the final results. The GAN loss is applied to the orthogonal projection on the subspace  $W$ .

**Conditional GAN loss** The LSGAN [Mao *et al.*, 2017] is applied as basic GAN loss because it brings smooth gradients which is more suitable for incremental training than vanilla GAN. Based on the relativistic discrimination used by the ESRGAN, we further lower the difficulty of discrimination for the orthogonal projections by introducing the conditional LR input. The input LR images will be bicubic-upsampled and concatenated with HR or SR images which construct  $\{x_W^h, x_V^c\}$  and  $\{x_W^s, x_V^c\}$ , and sent to the discriminator.

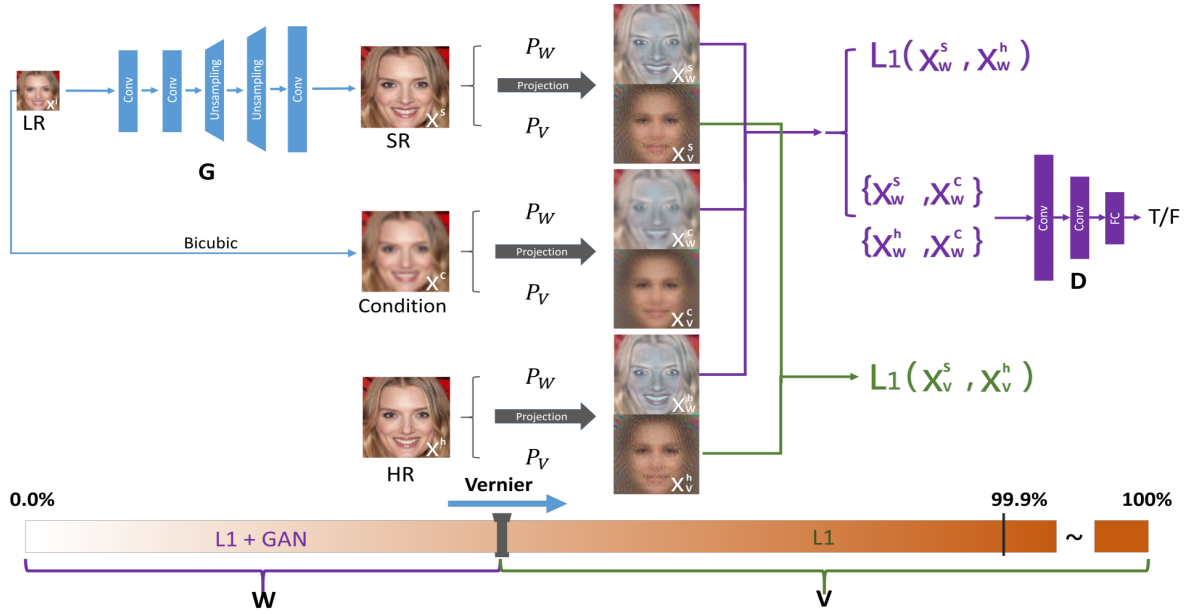


Figure 3: The pipeline of the PCA-SRGAN. The progress bar reveals the changing process of subspaces, above which other part illustrates the discrimination or constraint for orthogonal projections on the subspaces.

We formulate the loss of generator as:

$$L_G = E_{x_W^s} [(D(\{x_W^s, x_W^c\}) - E_{x_W^h} [D(\{x_W^h, x_W^c\})] - 1)^2] + E_{x_W^h} [(D(\{x_W^h, x_W^c\}) - E_{x_W^s} [D(\{x_W^s, x_W^c\})])^2], \quad (4)$$

and the loss of discriminator is the symmetrical form :

$$L_D = E_{x_W^h} [(D(\{x_W^h, x_W^c\}) - E_{x_W^s} [D(\{x_W^s, x_W^c\})] - 1)^2] + E_{x_W^s} [(D(\{x_W^s, x_W^c\}) - E_{x_W^h} [D(\{x_W^h, x_W^c\})])^2]. \quad (5)$$

The final GAN loss is:

$$L_{GAN}(G, D) = L_G + L_D. \quad (6)$$

**pixel-constrained L1 loss** We apply the 1-norm distance for the orthogonal projections on two subspaces between SR images and HR images.  $\alpha$  implies the different strength of constraint placed on the two orthogonal projections.

$$L_1(G) = E_{x_W^s} [\|x_W^s - x_W^h\|_1] + \alpha E_{x_V^s} [\|x_V^s - x_V^h\|_1]. \quad (7)$$

**Total loss** We conclude the total loss as

$$L_{Total}(G, D) = L_1(G) + \beta L_{GAN}(G, D), \quad (8)$$

where the weight  $\beta$  is used to balance different loss terms.

### 3.4 Algorithm summary

Let  $\phi$  denote the training set which contains  $m$  samples of LR face images  $\{x_1^l, x_2^l, \dots, x_m^l\}$  and HR face images  $\{x_1^h, x_2^h, \dots, x_m^h\}$ .  $\bar{x}$  denotes the average face of all HR face images in the training set.  $d$  denotes the dimension of the full space while  $d'$  denotes the dimension of subspace  $W$  when it increase to 99.9% of the full space. The whole algorithm has been summarized in *Algorithm 1*.

---

#### Algorithm 1 Incremental orthogonal projection discrimination.

---

**Input:**

$\phi, P, d, d', \bar{x}, G, D$ ; The step size  $k$  for each epoch.

**Output:**

$G$ .

- 1: Initialize the dimension of subspace  $W$ :  $n = 0$ .
  - 2: **while**  $n \leq d'$  **do**
  - 3: Start the new epoch:  $n = n + k$ .
  - 4: Separate the principal component vectors to construct orthogonal complementary subspace  $W$  and  $V$ :  
 $P_W = \{p_1, p_2, \dots, p_n\}$ ;  
 $P_V = \{p_{n+1}, p_{n+2}, \dots, p_d\}$ .
  - 5: **for** each batch of samples  $\{x^l, x^h\}$  in  $\phi$  **do**
  - 6: Generate the super-resolved face images:  
 $x^s = G(x^l)$ .
  - 7: Take the bicubic-upsampled  $x^l$  as condition  $x^c$ .
  - 8: Subtract the average face:  
 $x^s \leftarrow x^s - \bar{x}$ ,  $x^c \leftarrow x^c - \bar{x}$ ,  $x^h \leftarrow x^h - \bar{x}$ .
  - 9: Get the orthogonal projections of  $x^s, x^c$  and  $x^h$  on the subspace  $W$ :  
 $x_W^s = P_W x^s, x_W^c = P_W x^c, x_W^h = P_W x^h$ ,
  - 10: Get the residues on complementary subspace  $V$ :  
 $x_V^s = P_V x^s, x_V^c = P_V x^c, x_V^h = P_V x^h$ ,
  - 11: Apply loss according to Eq 8.
  - 12: Back-propagation and update  $G$  and  $D$ .
  - 13: **end for**
  - 14: **end while**
  - 15: **return**  $G$ ;
-

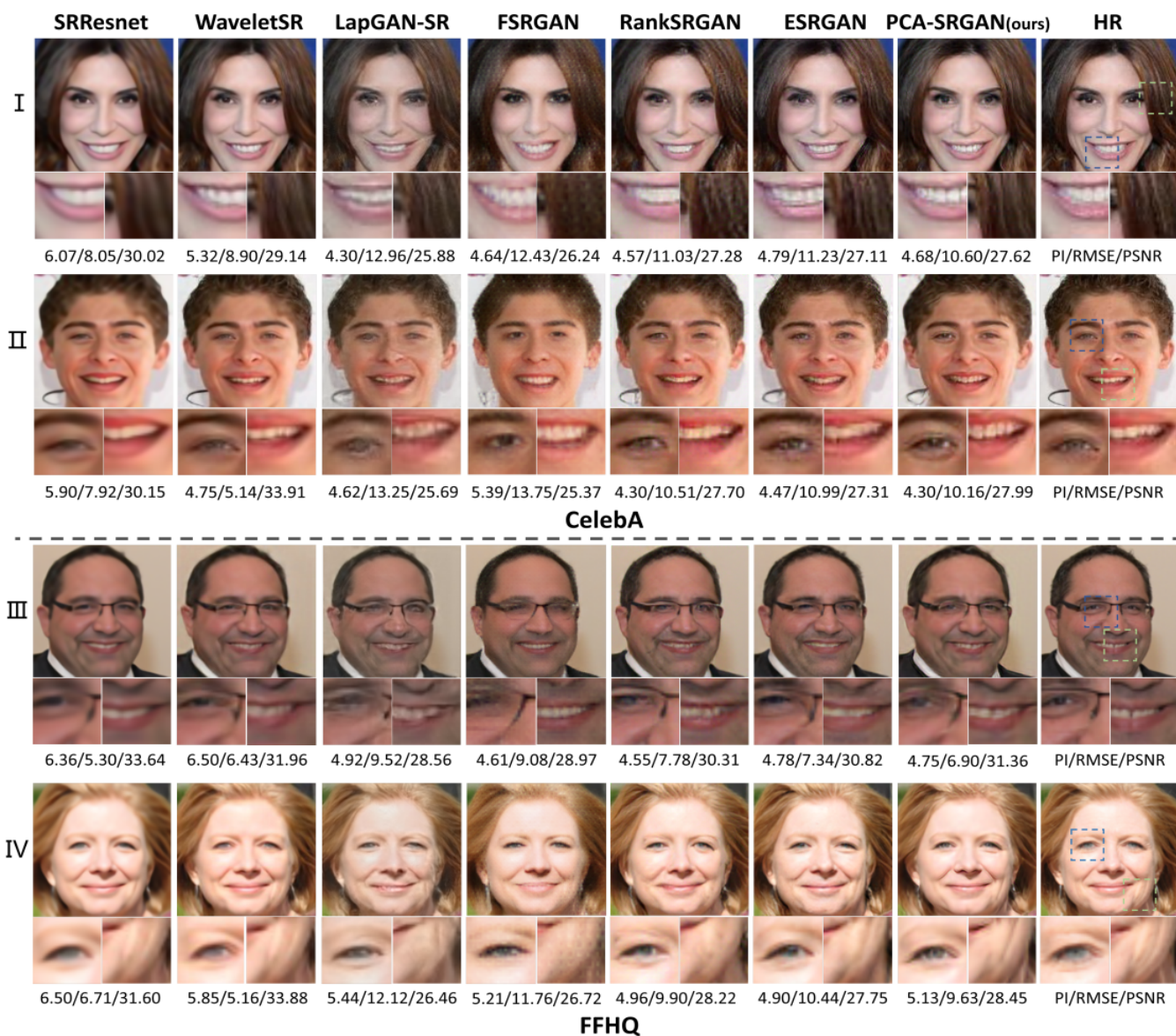


Figure 4: Comparisons of related methods on two datasets.

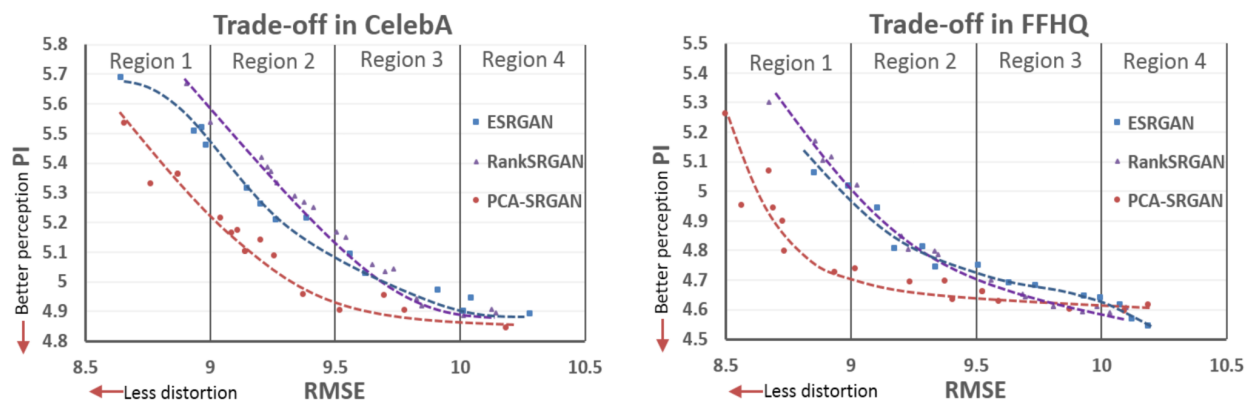


Figure 5: The trade-off boundaries between perception and distortion for ESRGAN, RankSRGAN and our PCA-SRGAN on two datasets.

Method	CelebA				FFHQ			
	Region 1	Region 2	Region 3	Region 4	Region 1	Region 2	Region 3	Region 4
ESRGAN	5.46/8.98/29.95	5.21/9.26/29.65	4.97/9.91/29.18	4.90/10.01/29.07	5.02/8.99/29.69	4.74/9.34/29.31	4.64/9.99/28.70	4.54/10.19/28.59
RankSRGAN	5.54/9.00/29.86	5.25/9.41/29.49	4.92/9.85/29.11	4.89/10.01/28.90	5.10/8.89/29.63	4.79/9.35/29.28	4.59/9.93/28.75	4.59/10.03/28.65
PCA-SRGAN	5.33/8.76/29.96	4.96/9.37/29.38	4.90/9.52/29.04	4.84/10.19/28.71	4.73/8.93/29.66	4.63/9.40/29.34	4.60/9.87/28.77	4.60/10.10/28.60
Groundtruth	4.88/0/-				4.68/0/-			

Table 1: The values of PI-RMSE and additional PSNR for compared works are provided in preset regions. We highlight the best performance with red and the second with blue.

## 4 Experiments

### 4.1 Datasets and preprocess

We have conducted the experiments on two face datasets.

**CelebA** CelebA consists of a large amount of celebrity face images cropped from the websites [Liu *et al.*, 2015]. Face images on CelebA are close to images taken in the real scene and contain rich texture, which brings realistic perceptual effect.

**FFHQ** FFHQ is a high-quality human face image dataset [Karras *et al.*, 2019] released recently and contains considerable variation in terms of age, ethnicity and background. Compared to CelebA, face images in this dataset is more clear and smooth with less noise.

Similar to FSRGAN, we select 18000 cropped and aligned face images with the size of  $128 \times 128$  as training sets and additional 100 images are used as validation sets for both CelebA and FFHQ. PCA operations are applied to the training sets of these two datasets. We get the mean and projection matrix for orthogonal projection discrimination on corresponding subspace.

### 4.2 Implementation details

Following SRGAN and ESRGAN, all experiments are performed with a scaling factor of  $4 \times$  between LR and HR images. The PSNR-oriented models are first obtained by pre-training on the corresponding face datasets. Then we employ the pre-trained models as an initialization for the generator and the Adam [Kingma and others, 2014] as the optimizer.

The training process contains 200 epochs divided into two stages according to the dimension of subspace  $W$ . In the first 100 epochs, the subspace grow up to 99% of the full space by increasing the dimension evenly and we set the learning rate as 0.0002 with loss weights  $\alpha = 1.0$  and  $\beta = 0.02$ . In next 100 epochs, the subspace grow up from 99% to 99.9% along with a linear decay for learning rate from 0.0002 to 0.00001, and the weights are set as  $\alpha = 0.0$  and  $\beta = 0.05$  to enlarge the effect of GAN loss and avoid the over-smooth of L1 loss. With this weight setting of loss functions, our model gets excellent visual effect as shown in Figure 4. Furthermore, the trade-off between perception and distortion of our model can be explored by applying a series of different weight settings.

### 4.3 Evaluation

We evaluate our PCA-SRGAN by qualitative visual effects and quantitative performance of the trade-off between perception and distortion. The PI and RMSE used in the PIRM2018 Challenge are chosen as the quantitative indices for perception and distortion respectively. PI is a no-reference image

quality measure and a lower value indicates better perceptual performance. RMSE is similar to the PSNR and positively correlated with the distortion degree.

**Qualitative results** We make comparisons with related methods, among which FSRGAN, ESRGAN and RankSRGAN have the aid of the perceptual (VGG) loss. As we can see in Figure 4, examples are displayed for visual comparison and the values of PI and MSE are provided as references. Considering the visual effect of whole faces, our PCA-SRGAN has outperformed previous approaches, reflected in more natural perception, clear texture and realistic facial components. For instance, PCA-SRGAN has produced visually pleasing sharper contour and facial components in all displayed examples than SRResnet which outputs the blurry eyes, teeth or hair and WaveletSR which generates some texture but still results in the smooth and unclear edge. PCA-SRGAN obtains less distortion and generates desired texture which is closer to groundtruth than other GAN-based methods, including those perceptual loss aided models. LapGAN-SR brings unnatural texture containing unpleasing noise (see Face II and III) which enlarges the distortion. FSRGAN produces warped face obviously in all face images which is likely to be caused by inaccurate facial prior prediction. ESRGAN and RankSRGAN achieve good perceptual results but our model obtains better visual effect on local patches, illustrated by clearer shape of the teeth (see Face I, II, and III), more life-like details in the eyes (see Face II and IV) as well as the vivid texture of hair (see Face I) and cheek (see Face IV). The superior visual performance has also approximately agreed with the referenced values below the images.

**Quantitative performance** Among all contrasted methods, we mainly seek and compare the boundaries of the PI-RMSE trade-off for three models — RankerGAN, ESRGAN and our PCA-SRGAN, which have shown obviously better visual effect in Figure 4 than other models. We choose a number of network weights that yield the PI values on the border over a certain range of RMSE which has been divided into four regions following the rules of the PIRM2018 Challenge. The PI and RMSE values construct the plotted points and we draw a trend line along the boundaries in the Figure 5. For each model the best PI value in each region is also selected in the Table 1 for comparison. In general, our model has achieved the superior performance of trade-off than ESRGAN and RankerGAN. For CelebA dataset, PCA-SRGAN overwhelms them by a much better trade-off boundary on whole regions. For FFHQ dataset, PCA-SRGAN has a comparable performance — ranking 1st on the first two regions and ranking 2nd on Region 3. It’s also reasonable for our model to have slight

weakness in Region 3 and 4 for the second dataset. Referring to the PI values of the groundtruth in Table 1, our model which aims to catch enough components of HR face has a convergence of perceptual index and would not produce the PI values far lower than the groundtruth boundlessly.

## 5 Ablation study

We have conducted experiments on the CelebA dataset as the ablation study. All of them share the same network structure, learning rate and training process.

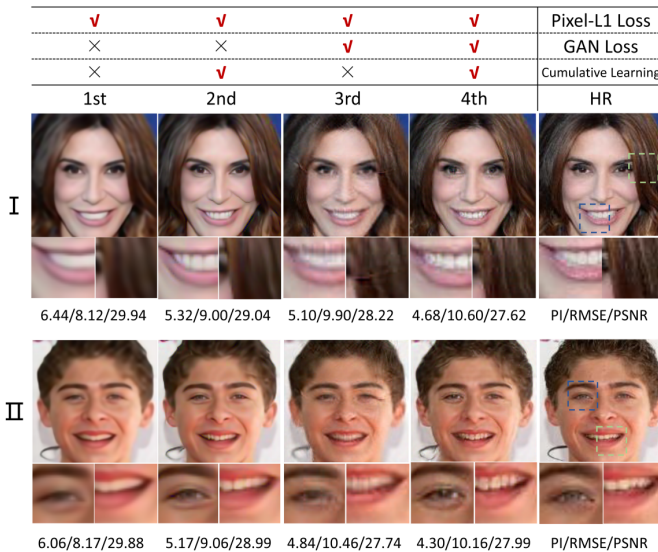


Figure 6: Overall comparisons for showing the effects of each part in PCA-SRGAN. Each column represents a model with its configurations in the top.

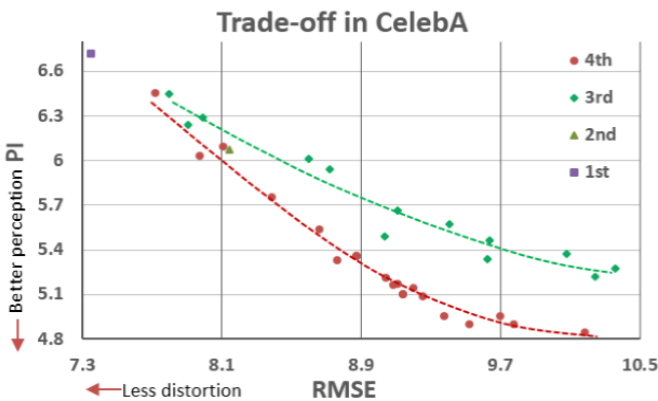


Figure 7: The comparison of quantitative Perception-Distortion trade-off. The GAN-based configurations can be plotted as the curves for the trade-off and other two configurations are plotted as single points in the figure.

The overall visual comparison is illustrated in Figure 6. Only with the Pixel-L1 loss (see the 1st column), the model generates smooth face without any texture. On the basis of

the Pixel-L1 loss, clear edge is produced by utilizing the cumulative learning of PCA projection (see the 2nd column). Combining the L1 loss with GAN loss which is applied to the fixed 99.9% PCA projection, the perceptual performance is improved but the GAN loss brings obvious noise and distortion (see the 3rd column). At last, under the same weights condition of GAN regularization, our model with cumulative learning obtains better visual performance and generates realistic texture (see the 4th column).

Furthermore, a quantitative comparison of Perception-Distortion trade-off for GAN-based configurations has been also given to validate the performance of cumulative learning strategy. Obviously as shown in Figure 7, the model with cumulative discrimination (the red curve) overwhelms the same model without cumulative strategy (the green curve) by a much better trade-off boundary on whole regions.

In a word, the cumulative learning of PCA projection along with the adversarial discrimination in our model has suggested convincing effectiveness to enhance the face super resolution task.

## 6 Conclusion

In this paper, a method named PCA-SRGAN using incremental orthogonal projection discrimination is proposed to enhance the performance of GAN on face SR task. We perform cumulative discrimination of orthogonal projections on PCA subspace to reduce the training difficulty and achieve precise and stable reconstruction without the help of perceptual (VGG) loss. Qualitative and quantitative comparisons have revealed the compelling performance of our model on super-resolving realistic face and the trade-off between perception and distortion. In the further research, we will develop our model by introducing other types of dictionary space or apply the model to other image processing tasks.

## References

- [Blau *et al.*, 2018] Yochai Blau, Roey Mechrez, et al. The 2018 pirm challenge on perceptual image super-resolution. In *Proceedings of the European Conference on Computer Vision*, 2018.
- [Bulat and Tzimiropoulos, 2018] Adrian Bulat and Georgios Tzimiropoulos. 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.
- [Chen *et al.*, 2018] Yu Chen, Ying Tai, et al. Fsrnet: End-to-end learning face super-resolution with facial priors. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018.
- [Denton *et al.*, 2015] Emily L Denton, Soumith Chintala, et al. Deep generative image models using a laplacian pyramid of adversarial networks. In *Advances in neural information processing systems*, 2015.
- [Dong *et al.*, 2014] Chao Dong, Chen Change Loy, et al. Image super-resolution using deep convolutional networks. *IEEE Trans Pattern Anal Mach Intell*, 2014.
- [Goodfellow *et al.*, 2014] Ian Goodfellow, Jean Pouget-Abadie, et al. Generative adversarial nets. In *International Conference on Neural Information Processing Systems*, 2014.
- [Huang *et al.*, 2017] Huaibo Huang, Ran He, et al. Waveletsrnet: A wavelet-based cnn for multi-scale face super resolution. In *Proceedings of the IEEE International Conference on Computer Vision*, 2017.
- [Johnson *et al.*, 2016] Justin Johnson, Alexandre Alahi, and Li Fei-Fei. Perceptual losses for real-time style transfer and super-resolution. In *European conference on computer vision*, 2016.
- [Jolliffe, 1986] I. T Jolliffe. *Principal component analysis*. 1986.
- [Karras *et al.*, 2017] Tero Karras, Timo Aila, et al. Progressive growing of gans for improved quality, stability, and variation. *arXiv preprint*, 2017.
- [Karras *et al.*, 2019] Tero Karras, Samuli Laine, and Timo Aila. A style-based generator architecture for generative adversarial networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2019.
- [Kim and Lee, 2018] Jun-Hyuk Kim and Jong-Seok Lee. Deep residual network with enhanced upscaling module for super-resolution. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2018.
- [Kim *et al.*, 2019] Deokyun Kim, Minseon Kim, et al. Progressive face super-resolution via attention to facial landmark. *arXiv preprint*, 2019.
- [Kingma and others, 2014] Diederik P Kingma et al. Adam: A method for stochastic optimization. *Computer Science*, 2014.
- [Lai *et al.*, 2017] Wei-Sheng Lai, Jia-Bin Huang, et al. Deep laplacian pyramid networks for fast and accurate super-resolution. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017.
- [Ledig *et al.*, 2017] Christian Ledig, Lucas Theis, 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.
- [Lim *et al.*, 2017] Bee Lim, Sanghyun Son, et al. Enhanced deep residual networks for single image super-resolution. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, 2017.
- [Liu *et al.*, 2015] Ziwei Liu, Ping Luo, et al. Deep learning face attributes in the wild. In *Proceedings of the IEEE international conference on computer vision*, 2015.
- [Mao *et al.*, 2017] Xudong Mao, Qing Li, et al. Least squares generative adversarial networks. In *International Conference on Computer Vision*, 2017.
- [Song *et al.*, 2017] Yibing Song, Jiawei Zhang, et al. Learning to hallucinate face images via component generation and enhancement. In *Twenty-Sixth International Joint Conference on Artificial Intelligence*, 2017.
- [Vasu *et al.*, 2018] Subeesh Vasu, Nimisha Thekke Madam, and AN Rajagopalan. Analyzing perception-distortion trade-off using enhanced perceptual super-resolution network. In *Proceedings of the European Conference on Computer Vision*, 2018.
- [Wang and Tang, 2005] Xiaogang Wang and Xiaoou Tang. Hallucinating face by eigentransformation. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 2005.
- [Wang *et al.*, 2018a] Xintao Wang, Ke Yu, et al. Esrgan: Enhanced super-resolution generative adversarial networks. In *Proceedings of the European Conference on Computer Vision*, 2018.
- [Wang *et al.*, 2018b] Yifan Wang, Federico Perazzi, et al. A fully progressive approach to single-image super-resolution. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2018.
- [Yang *et al.*, 2018] Chih-Yuan Yang, Sifei Liu, et al. Hallucinating compressed face images. *International Journal of Computer Vision*, 2018.
- [Yu *et al.*, 2018] Xin Yu, Basura Fernando, et al. Super-resolving very low-resolution face images with supplementary attributes. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018.
- [Zhang *et al.*, 2019] Wenlong Zhang, Yihao Liu, et al. Ranksgan: Generative adversarial networks with ranker for image super-resolution. In *Proceedings of the IEEE International Conference on Computer Vision*, 2019.