改进的人脸超分辨率生成对抗网络

王梦雪[1]，陈振学[2], 周新洁[3]

[1] 山东大学控制科学与工程学院，济南 250061

[2] 山东大学控制科学与工程学院，济南 250061

[3] 山东大学深圳研究院，深圳，518057

摘要：人脸超分辨率方法用于从低分辨率图像生成高分辨率图像，以便更好地可视化。超分辨率生成对抗网络是一项具有开创性的工作, 它可以生成具有逼真纹理的单个超分辨率图像。基于超分辨生成对抗网络，本文提出了改进的人脸超分辨率生成对抗网络。本文深入研究了超分辨生成对抗网络架构的关键组成部分，并对每个部分进行了改进，以实现更强大的超分辨生成对抗网络。首先，超分辨生成对抗网络使用残余块作为深度生成器网络的核心, 本文决定采用密集卷积网络块（密集块）作为深度生成器网络。此外，尽管生成对抗网络性能优越，但难以训练。本文使用一种简单有效的正则化方法 -谱归一化生成对抗网络来解决该问题。通过实验证实，本文提出的方法在训练稳定性和视觉改善方面优于其他现有方法。

关键词：人脸超分辨，生成对抗网络，谱归一化，密集块

中图分类号： TP391

Improved Face Super-Resolution Generative Adversarial Networks

Wang Mengxue[1], Chen Zhenxue[2], Zhou Xinjie[3]

[1] School of Control Science and Engineering, Shandong University, Jinan 250061

[2] School of Control Science and Engineering, Shandong University, Jinan 250061

[3] Shenzhen Research Institute of Shandong University, Shenzhen 518057

Abstract: The face super-resolution method is used for generating high-resolution images from low-resolution ones for better visualization. The Super-Resolution Generative Adversarial Network (SRGAN) can generate a single super-resolution image with realistic textures, which is a groundbreaking work. Based on SRGAN, we propose improved face super-resolution generative adversarial networks. The super-resolution image details generated by SRGAN usually have undesirable artifacts. To further improve visual quality, we

Foundations: National Natural Science Foundation of China (61203261 and 61876099)，Shenzhen Science and Technology Research and Development Funds (JCYJ20170307093018753)，Shenzhen Science and Technology Research and Development Funds (JCYJ20180305164401921)

Author Introduction: Wang Mengxue(1994-)，female，master candidate，major research direction：face super-resolutiom, low-resolution face recognition, deep learning, computer vision and machine learning，e-mail: 201714400@mail.sdu.edu.cn. Correspondence author：Chen Zhenxue(1977-)，male，professor，major research direction：image processing, pattern recognition, and computer vision, with applications to face recognition, e-mail: chenzhenxue@sdu.edu.cn. Zhou Xinjie(1994-),female, postgraduate，major research direction：face recognition, e-mail: 2990924993@qq.com.

delve into the key components of the SRGAN network architecture and improve each part to achieve a more powerful SRGAN. First, the SRGAN employs residual blocks as the core of the very deep generator network G. In this paper, we decide to employ Dense Convolutional Network blocks (Dense blocks), which connect each layer to every other layer in a feed-forward fashion as our very deep generator networks. Moreover, in the past few years, generative adversarial networks (GANs) have been applied to solve various problems. Despite its superior performance, however, it is difficult to train. A simple and effective regularization method called spectral normalization GAN (SNGAN) is used to solve this problem. We have experimentally confirmed that our proposed method is superior to the other existing method in training stability and visual improvements.

Key words: Face super-resolution, GAN, spectral normalization, dense blocks.

0 Introduction

Face Super-Resolution (SR), also known as face hallucination, aims to generate a High-Resolution (HR) face image from a Low-Resolution (LR) input. Since Dong et al.[1] proposed SRCNN, various deep convolutional neural network methods have been proposed to improve SR performance, especially to improve peak signal-to-noise ratio (PSNR) values. However, a higher PSNR value does not mean that the visual quality is better. They tend to output excessively smooth results while ignoring high frequency details. We are more eager to get higher quality images rather than images with higher PSNR values. SRGAN[2] is one of the milestones in the pursuit of visually pleasing results. However, there is still a significant gap between the SRGAN results and the real image. In this paper, we study the key components of SRGAN and improve the model from three aspects.

First, the discriminator is improved by using spectral normalization GAN (SNGAN)[3]. GAN has recently attracted great interest and has gradually improved the framework and its many applications. However, during training, the discriminator is usually unstable. In[4], Arjovsky et al. propose a method called Wasserstein-GAN (WGAN). And then Gulrajani et al.[5] propose an improved WGAN model with gradient penalty to improve the stability and performance of WGAN. Miyato et al.[3] provide a simpler normalization method – spectral normalization as a stabilizer of training of GANs. Inspired by this, spectral normalization is employed on our discriminator network, which is effective and has small additional computational cost.

Second, the dense blocks which have higher capacity and are easier to train are introduced to improve the network structure. We remove the Batch Normalization (BN)[6] layers as in[7]. In recent years, as convolution neural networks (CNNs) have become increasingly deeper, the problem of vanishing-gradient has also emerged. In addition, in the previous works[1,8], only the high-level features are used to reconstruct HR images. The low-level potential features which can provide additional information are ignored. Using a skip connection to create a short path from the top to the bottom is a good way to solve this problem. This helps to convey information and gradients over the network, making it easy to train. A new structure called dense convolutional network (DenseNet)[9] is proposed, in which any layer of each block is the output of all the previous layers, as well as the input of all subsequent layers. In SRGAN[2], the authors choose to use the residual block as the core of the very deep generator network G. In this paper, we employ the more advanced dense blocks without BN layers as our core of the generator network. The benefit of using it is expected to be better quality reconstructed images.

We summarize our contribution as follows:

(I) To the best of our knowledge, we systematically studied the application of the newly proposed spectral normalization GAN on the single face image super-resolution problem for the first time.

(II) We demonstrate that the generator network with the dense blocks without BN layers as basic blocks can achieve good reconstruction performance and that the reconstruction performance of the generator network can be further improved by the fusion of features at different levels through dense skip connections.

This paper is organized as follows. The related works are described in Section II. Our proposed method is presented in Section III. The loss functions are explained in Section IV. Our experiments and results are described in Section V. Finally, our conclusions are put forth in Section VI.

1 Related works

This section reviews related works in image and face super-resolution, and generative adversarial networks.

1.1 Image super-resolution

There are many single image super resolution (SISR) methods already in existence. Among them, the interpolation method is the simplest and the most widely used method. Timofte et al.[10] and Yang et al.[11] introduce the sparse-based techniques to enhance linear models with rich image priors.

Recently, many image super-resolution methods based on deep neural networks have been proposed[1,8,12-17]. The first application of convolutional neural networks to image super-resolution problems is the SRCNN[1] proposed by Dong et al. Later, Shi et al.[12] propose an efficient subpixel convolutional neural network (ESPCN). Dong et al.[8] develop an accelerated superresolution convolutional neural network. Kim et al.[13] propose a very deep CNN using the residual network to solve the super-resolution problem (VDSR). Lai et al.[14] propose the Laplacian pyramid super-resolution Network (LapSRN) to solve the speed and accuracy of the SR problem by taking the original LR image as input and gradually reconstructing the sub-band residuals of the HR images. Tai et al.[15] present the Deep Recursive Residual Network (DRRN) to address the problems of model parameters and accuracy. Moreover, Ledig et al.[2] propose the Super-Resolution Generative Adversarial Network (SRGAN) for photo-realistic image SR using a perceptual loss function that consists of an adversarial loss and a content loss.

1.2 Face super-resolution

Baker et al.[18] develop a face hallucination method using Bayesian formulation. Liu et al.[19] present a two-step approach to hallucinate faces. Wang et al.[20] propose a face hallucination method by Eigen transformation. Yang et al.[21] present a face hallucination method by exploiting the local structure. Jin et al.[22] propose a robust multi-image based blind face hallucination. Huang et al.[23] present a multi-scale face super resolution method based on the wavelet-based CNN.

1.3 GANs

Although GAN has great potential and popularity, its performance is still affected by several problems. Arjovsky et al.[24] propose replacing the original loss function with a new loss function based on the Wasserstein distance. Gulrajani et al.[5] propose an improved WGAN model that uses gradient penalty instead of weight clipping. To further stabilize the training of discriminator, a more advanced weight normalization technique called spectral normalization is proposed by Miyato et al.[3]. It has been shown that spectral normalization can improve the sheer quality of the generated images better than weight normalization[25] and gradient penalty[1]. In this work, we improve SRGAN by employing a more effective spectral normalization GAN.

2 Proposed Method

Our main aim is to produce the more realistic super-resolved images. In this section, we describe the proposed model architecture. It consists of two parts: the first part is a generator network used to super-resolve the LR images (I^{LR}). The second part is a discriminator used to distinguish between the super-resolved (I^{SR}) and the original HR images (I^{HR}).

2.1 Network Structure

For the generator network, in order to improve the recovered image quality of SRGAN, inspired by[9], unlike the residual blocks used in the original SRGAN, dense blocks without the BN layers are used in our paper. In the discriminator network, the spectral normalization is used to stabilize the training of the discriminator. We will provide the intuition and motivation behind our design choices. Their differences will be detailed below.

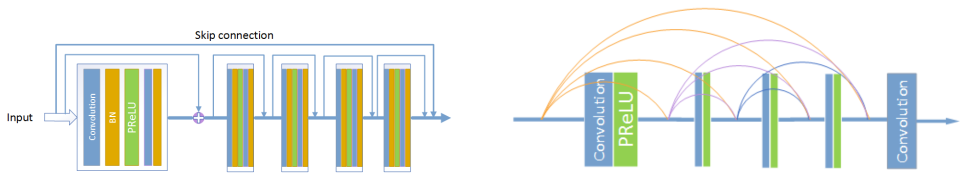
2.1.1 Generator

In the experiment, we observe that when the network is deeper and trained under the GAN framework, the BN layer may introduce the undesired artifacts. These artifacts occasionally occur in iterations and different settings, resulting in an unstable training. Therefore, in order to ensure the stability and consistency of the training, we remove the BN layers in generator network.

The generator architecture of[2] uses 16 residual blocks with identical layout. The structure of the residual blocks is shown in Fig. 1(a). We propose to use dense blocks. Our motivation behind this change is as follows: since the main purpose of this network is to generate super-resolution images, using the normal residual blocks is inadequate for the generation of details.

The SR-ResNet of[2] combines 16 residual blocks into a large block, with a skip connection that links the first and the last block, in an attempt to improve the gradient flow. It may not fully explore the advantages of the skip connections. We argue that the resolution is to use a new network - DenseNet. In the structure of DenseNet, a short path is created between the layer and each of the other layers. This strengthens the flow of information through deep networks. In addition, because DenseNet can significantly reduce the number of parameters through feature reuse, it requires less memory and computation for high performance. Here, the dense blocks are used as the basic building blocks in our generator network. The structure of each dense block can be seen in Fig. 1(b). Specifically, in our paper, we use each dense block with 16 convolutional layers. If each convolution layer produces k feature maps as output, the total number of feature maps generated by one dense block is $k \times 16$, where k is called growth rate. The growth rate k adjusts how much new information per layer contributes to the final reconstruction. In order to prevent the network from being too wide, we only use one dense block and set the growth rate k in this paper to 12.

In the proposed generator network as shown in Fig. 2 (a), all of the feature maps generated by the network are used as inputs to subsequent sub-pixel layers. If a large number of feature maps are fed directly into the sub-pixel layers, the computational cost and model size are significantly increased. Therefore, a convolutional layer with a 1×1 kernel is used as a bottleneck layer in this paper to reduce the number of input feature maps. It has been proved in previous studies[26] that this method can effectively reduce the number of input feature maps and improve computational efficiency. We input the feature maps which obtained after using the bottleneck layer into the sub-pixel layers. We still use two trained sub-pixel convolutional layers[12] to increase the resolution of the input images, as done in SRGAN.



(a) The structure of the residual block. (b) The structure of the dense block.

图 1: The structure of the residual blocks and the dense blocks.

2.1.2 Discriminator

In addition to the improved generator structure, we have also enhanced the discriminator based on spectral normalization GAN. We train the discriminator network to distinguish the generated SR image from the original HR image. The architecture is shown in Fig. 2 (b). We generally follow the architecture proposed by SRGAN. It contains eight convolutional layers, each with a filter kernel of 3×3. The number of filter kernels increases from 64 to 512 by a rate of 2 times. Finally, in order to obtain the probability of sample classification, two dense layers and a sigmoid activation function are set. We use LeakyReLU activation ($\alpha = 0.2$) and avoid max-pooling throughout the network. We replace the standard discriminator with the spectral normalization discriminator, denoted as D_{SN}.The method we will employ in this paper, called spectral normalization, is a method that aims to skirt this issue by normalizing the weight matrices using the technique devised by[3]. The spectral normalization normalizes the spectral norm of the discriminator weight matrix W so that it satisfies the Lipschitz constraint $\sigma(W) = 1$:

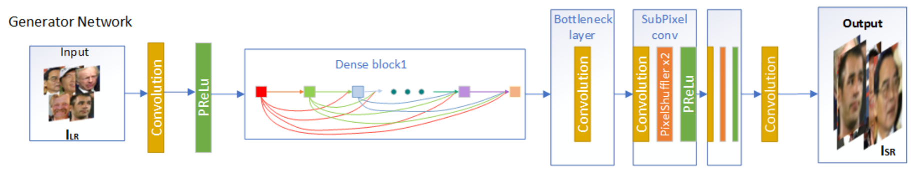
$$W_{SN}(W) = W/\sigma(W) \tag{1}$$

where $\sigma(W)$ is a singular value of matrix W.

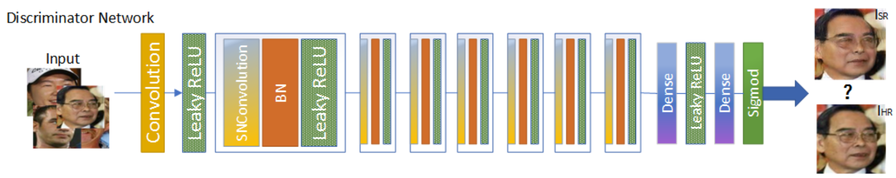
3 Loss Function

3.1 Content loss

Pixel-wise loss. Given a low resolution image I^{LR} and the corresponding high resolution image I^{HR}, the pixel-wise MSE loss is used to minimize the distance between the high resolution



(a) Architecture of Generator Network.



(b) Architecture of Discriminator Network.

图 2: Architecture of Generator and Discriminator Network

and the super-resolved image. It is defined as follows:

$$l_{pixel}^{SR} = \frac{1}{r^2WH}\sum_{x=1}^{rW}\sum_{y=1}^{rH}\left(I_{x,y}^{HR} - G\left(I^{LR}\right)_{x,y}\right)^2 \tag{2}$$

where W and H denote the size of I^{LR} and r is the upsampling factor (set to 4 in our case).

Perceptual Loss. Although the pixel-based MSE loss achieves a high PSNR value, it often results in images lacking high frequency content. In order to solve this problem, in[2,27], the perceptual loss is proposed. In our work, perceptual loss is defined on the ReLU active layer based on the pre-trained 19-layer VGG network described by Simonyan and Zisserman[28]. We then define the loss as the Euclidean distance between the feature representations of the reconstructed image and the ground truth:

$$l_{percep}^{SR} = \frac{1}{W_{i,j}H_{i,j}}\sum_{x=1}^{W_{i,j}}\sum_{y=1}^{H_{i,j}}\left(\phi_{i,j}\left(I^{HR}\right)_{x,y} - \phi_{i,j}\left(G\left(I^{HR}\right)\right)_{x,y}\right)^2 \tag{3}$$

where $W_{i,j}$ and $H_{i,j}$ describe the dimensions of the respective features maps within the VGG network, i represents the feature mapping obtained by the j-th convolution before the i-th largest pooling layer in the VGG19 network. We use the layer ReLU54, which gives good empirical in our experiments.

3.2 Adversarial loss

We choose to experiment with two adversarial losses separately. First is the following standard objective function:

$$V\left(G,D\right) = E_{I^{HR}\sim p_{train\left(I^{HR}\right)}}\left[logD\left(I^{HR}\right)\right] + E_{I^{LR}\sim p_{G\left(I^{LR}\right)}}\left[log\left(1-D\left(G\left(I^{LR}\right)\right)\right)\right] \tag{4}$$

The generative loss l_{Genx} is defined based on the probabilities of the discriminator D over all the training samples as:

$$l_{Gen1} = \sum_{n=1}^{N} -logD\left(G\left(l^{LR}\right)\right) \tag{5}$$

And then, We also test the performance of the algorithm with the so-called hinge loss proposed in Spectral Normalization GAN [3], which is given by:

$$V_D\left(G, D\right) = E_{I^{HR} \sim p_{train\left(I^{HR}\right)}}\left[min\left(0, -1 + D\left(I^{HR}\right)\right)\right] + E_{I^{LR} \sim p_{G\left(I^{LR}\right)}}\left[min\left(0, -1 - D\left(I^{LR}\right)\right)\right] \tag{6}$$

$$V_G\left(G, D\right) = -E_{I^{LR} \sim P_G\left(I^{LR}\right)}\left[D\left(G\left(I^{LR}\right)\right)\right] \tag{7}$$

respectively for the discriminator and the generator.

$$l_{Gen2} = \sum_{n=1}^{N} -D\left(G\left(I^{LR}\right)\right) \tag{8}$$

The algorithm based on the hinge loss also shows good performance when evaluated with PSNR and SSIM.

3.3 Overall training loss

The objective function l^{LR} for training our model is expressed as :

$$l^{SR} = l_X^{SR} + \lambda l_{Genx} \tag{9}$$

where λ is the corresponding weights, X is pixel or percep, $x = 1$ or 2.

4 Experiments

In this section, we introduce our experiments. First, we describe the datasets used in training and testing. Second, we explain our implementation details. Then, we provide the results compared with other methods. Finally we present our analysis and discussion.

4.1 Datasets and metrics

Datasets. During experiment, we need the high resolution images datasets and the low resolution image datasets. The low-resolution images are obtained by 4× downscaling of the high-resolution image. We conduct extensive experiments on the two datasets: LFW [29] and celebA [30]. We use the aligned and cropped celebA dataset as the training dataset. We only use all available data and do not need to group the face images into different pose and facial expression subcategories. Our network never sees the test LR images in the training phase.

Then we randomly select 1000 images of the LFW dataset as the test set. We coarsely crop the images according to their face regions and resize them to 128×128 without any pre-alignment operation. We use the MTCNN[31] to obtain the cropped image as the input.

Metrics. All experiments are performed with a scale factor of 4x between low- and high-resolution images. The peak signal-to-noise ratio (PSNR) and the structural similarity (SSIM) index were used as the metrics for evaluation. For fair comparison, all reported PSNR (dB) and SSIM measures are calculated on the y-channel of center-crooped.

4.2 Implementation Details

We train all networks on an NVIDIA GeForce GTX 1060 6GB using celebA database. For each mini-batch, we crop 16 random 96×96 sub images of distinct training images. We can apply the generator model to images of arbitrary size. The training process is divided into two parts. First, we train the MSE-based SRDenseNet network for high PSNR values. The SRDenseNet networks are trained with a learning rate of 10^{-4} and 10^{6} update iterations. Then, when training the actual GAN, we employ the trained MSE-based SRDenseNet network as initialization. The loss function in Eq. 9 is used to train the generator with $\lambda = 0.003$. All SNGAN variants are trained with 10^{5} update iterations at a learning rate of 10^{-4} and another 10^{5} iterations at a lower rate of 10^{-5}. Pre-training MSE-based SRDenseNet network helps GAN-based methods achieve more visually satisfying results. The reasons are as follows: I) for the generator, it can avoid undesired local optima. II) After pre-training, the discriminator initially receives the good super-resolution images instead of the extremely fake images, which helps it pay more attention to texture discrimination.

For optimization we use Adam with $\beta_1 = 0.9$. We alternate updates to the generator and discriminator network, which is equivalent to $k = 1$ as used in Goodfellow et al.[32]. Our generator has two settings - one is similar to SRGAN, which contains 16 residual blocks, and the other is a model with one dense block. Our discriminator has three settings - one is the standard GAN discriminator, another is the discriminator of Wasserstein GAN, and the last one is the discriminator of SNGAN.

The spectral norm $\sigma(w)$ that we use to regularize each layer of the discriminator is the largest singular value of W. We estimate $\sigma(w)$ by using the power iteration method[33-34]. In the experiment we find that one round of power iteration is adequate to achieve satisfactory performance. In comparison to the full computational cost of the standard GANs, this method is very computationally cheap.

4.3 Results

We compare our approach with two other types of approaches: (I) general super-resolution (SR) method; and (II) the face super-resolution method. For SR method, Kim et al. [13] propose a method called VDSR that uses a very deep convolutional network to obtain super-resolution generic images. Ledig et al. [2] propose a general-purpose super-resolution method that uses generative adversarial networks and trains with pixel-wise and adversarial losses called SRGAN. Ma et al. [35] reconstruct HR face images using location patches in the dataset. Chen et al. [36] propose using Wasserstein GANs for face super-resolution.

4.3.1 Qualitative Comparisons with SoA

Bicubic interpolation only upsamples the intensity of the image from neighboring pixels, rather than generating new content for new pixels. As shown in Fig. 3 (c), bicubic interpolation does not produce facial details.

VDSR uses only the pixel loss of l_2 in training. As shown in Fig. 3 (d), VDSR cannot generate real facial details, and super-resolution faces are still blurred.

SRGAN is able to take advantage of the adversarial loss to enhance the details. However, SRGAN does not make full use of the effective information and is difficult to train. And sometimes, SRGAN can produce artifacts such as adding wrinkles to the face, as shown in Fig. 3 (e).

Ma et al.'s approach [35] avoids ghosting artifacts caused by global models such as PCA because they employ a method of local constraints learned from positioned sample patches. However, it requires precise alignment of the exemplar patches. As shown in Fig. 3 (f), the results of this method have obvious blur artifacts due to the unaligned location patches in the datasets we are using.

Chen et al.'s method [36] for improvement is only about the stability and easiness training, not the quality of super-resolution facial images, as shown in Fig. 3 (g).

Our model is capable of making full use of the effective information of each layer of neural networks and removing the undesired artifacts, as shown in Fig. 3 (g).

For a fair comparison, we use the released codes of the above model and use the same dataset CelebA for training.

4.3.2 Quantitative Comparisons with SoA

We also assess the performance of all methods quantitatively by comparing the average PSNR and the structural similarity (SSIM) scores on the entire test dataset. TABLE I indicates that our method achieves better performance compared to other methods. Our method achieves facial details consistent with real faces because it attains the best PSNR and SSIM results.

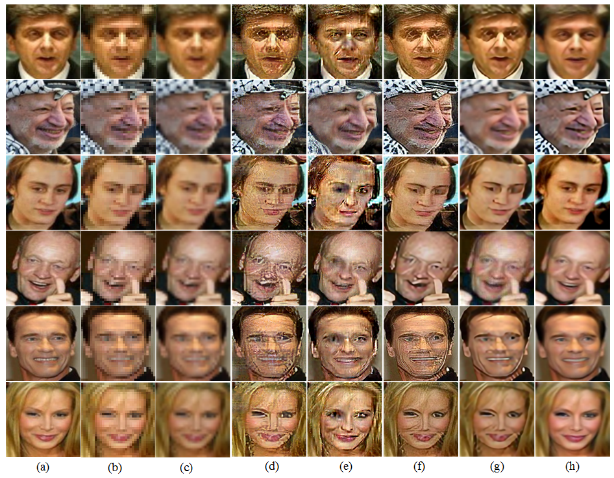


图 3: Comparisons with the state-of-art methods. (a) Original HR images. (b) LR inputs. (c) Bicubic interpolation. (d) Kim et al's method[13] (VDSR). (e) Ledig et al.'s method[2] (SRGAN). (f) Ma et al.'s method[35]. (g) Chen et al.' s method[36]. (h) Our method

表 1: Quantitative comparisons on the entire test dataset

Methods	Bicubic	VDSR[13]	SRGAN[2]	Ma[35]	Chen[36]	Ours
PSNR	28.25	28.10	27.79	28.14	27.93	**29.03**
SSIM	0.80	0.79	0.76	0.78	0.76	**0.84**

4.4 Analysis and Discussion

We investigate the effects of four important components in our framework:

Effects of removing BN layers. In order to explore the impact of the BN layers, we conduct several experiments. All BN layers are before the upsampling layer. As shown in Fig. 4, we demonstrate that visual results without BN layers can achieve stable and consistent performance without artifacts. It saves computing resources and memory usage without compromising performance.

Effects of the SNGAN. In our framework, we use a novel weight normalization method called spectral normalization. It can stabilize the training of discriminator networks and is simple to implement. As shown in Fig. 5, we explore the results with different weight normalization methods.

Effects of the dense blocks. The dense blocks are the core of our framework. In the comparison experiment, the other components are identical except for dense blocks. As indicated in Fig. 6, the restored texture can be further improved using the proposed dense blocks.

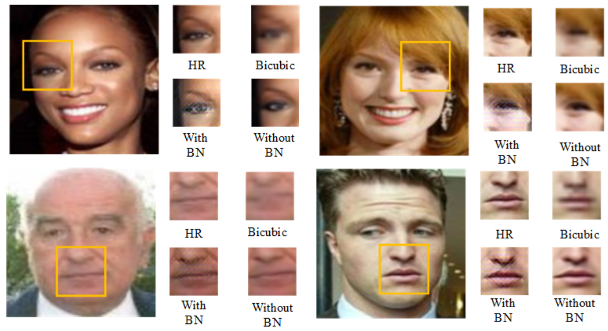


图 4: Ablation study of BN removal

图 5: Ablation study of GAN. (a) Original HR images. (b) standard GAN. (c) Wasserstein GAN. (d) SNGAN

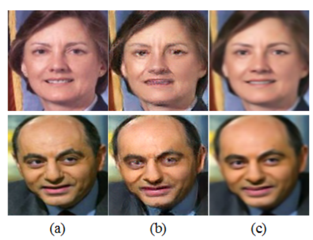


图 6: Ablation study of Dense block. (a) Original HR images. (b)Residual block. (c) Dense block

5 Conclusion

We have presented an effective method to super-resolve the LR face images by exploiting a generative adversarial network. First, we propose to use more advanced dense blocks without BN layers instead of residual blocks. Dense blocks can get more and more effective information through dense connections. We then further introduce the use of spectral normalization GAN as a stabilizer of training of GANs. When we apply spectral normalization to the GANs on face super-resolution tasks, it can achieve better results relative to previous works. Moreover, we add the hinge loss in our experiment, which can recover more detailed texture s.

参考文献（References）

[1] Dong C, Loy C C, He K, et al. Image Super-Resolution Using Deep Convolutional Networks[J]. IEEE Trans Pattern Anal Mach Intell, 2016, 38(2):295-307.

[2] Ledig C, Theis L, Huszar F, et al. Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network[J]. 2016.

[3] Miyato T, Kataoka T, Koyama M, et al. Spectral Normalization for Generative Adversarial Networks[J]. arXiv preprint arXiv:1802.05957, 2018

[4] Arjovsky M, Chintala S, Bottou L. Wassterstein GAN[J]. arXiv preprint arXiv:1701.07875, 2017

[5] Gulrajani I, Ahmed F, Arjovsky M, et al. Improved Training of Wasserstein GANs[C]. Advances in Neural Information Processing Systems. 2017: 5767–5777

[6] Ioffe S, Szegedy C. Batch normalization: Accelerating deep network training by reducing internal covariate shift[C]. ICML. 2015.

[7] Tong T, Li G, Liu X, et al. Image Super-Resolution Using Dense Skip Connections[C]. IEEE International Conference on Computer Vision (ICCV). IEEE Computer Society, 2017.

[8] Dong C, Loy C C, Tang X. Accelerating the super-resolution convolutional neural network[C]. European conference on computer vision. : Springer, 2016: 391–407.

[9] Gao H, Zhuang L, Maaten L V D, et al. Densely connected convolutional networks [C]. IEEE Conference on Computer Vision and Pattern Recognition. 2017.

[10] Timofte R, Vincent De Smet, Luc Van Gool. A+: Adjusted Anchored Neighborhood Regression for Fast Super-Resolution[C]. Asian conference on computer vision. : Springer, 2014: 111–126.

[11] Yang J, Wright J, Huang T, et al. Image super-resolution as sparse representation of raw image patches[C]. IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2008.

[12] Shi W, Caballero J, HUSZ¨¢R F, et al. Real-time single image and video superresolution using an efficient sub-pixel convolutional neural network[C]. IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2016.

[13] Kim J, Lee J K, Lee K M. Accurate Image Super-Resolution Using Very Deep Convolutional Networks[C]. IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2016.

[14] Lai W S, Huang J B, Ahuja N, et al. [IEEE 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) - Honolulu, HI, USA (2017.7.21-2017.7.26)] 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) - Deep Laplacian Pyramid Networks for Fast and Accurate Super-Resolution[J]. 2017:5835-5843.

[15] Ying T, Jian Y, Liu X. Image Super-Resolution via Deep Recursive Residual Network[C]. IEEE Conference on Computer Vision and Pattern Recognition. 2017.

[16] Mao X J, Shen C, Yang Y B. Image Restoration Using Very Deep Convolutional Encoder-Decoder Networks with Symmetric Skip Connections[J]. arXiv preprint arXiv:1603.09056, 2016, 2.

[17] Kim J, Lee J K, Lee K M. Deeply-Recursive Convolutional Network for Image Super-Resolution[C]. Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 1637–1645.

[18] Baker S, Kanade T. Hallucinating Faces[C]. IEEE International Conference on Automatic Face and Gesture Recognition. 2000.

[19] Liu C . A two-step approach to hallucinating faces : Global parametric model and local non-parametric model[C]. IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Hawaii, USA, Dec. 2001. IEEE, 2001.

[20] Wang X , Tang X . Hallucinating Face by Eigentransformation[J]. IEEE Transactions on Systems, Man and Cybernetics, Part C (Applications and Reviews), 2005, 35(3):425-434.

[21] Yang C Y, Liu S, Yang M H. Structured Face Hallucination[C]. IEEE Conference on Computer Vision and Pattern Recognition. 2013.

[22] Jin Y, Bouganis C S. Robust multi-image based blind face hallucination[C]. IEEE Conference on Computer Vision and Pattern Recognition. 2015.

[23] Huang H, He R, Sun Z, et al. Wavelet-SRNet: A Wavelet-Based CNN for Multi-scale Face Super Resolution[C]. 2017 IEEE International Conference on Computer Vision (ICCV). IEEE Computer Society, 2017.

[24] Arjovsky M , Bottou, L. Towards Principled Methods for Training Generative Adversarial Networks[J]. Stat, 2017.

[25] Salimans T, Kingma D P. Weight Normalization: A Simple Reparameterization to Accelerate Training of Deep Neural Networks[C]. AdvancesinNeuralInformationProcessing Systems. 2016: 901–909.

[26] Szegedy C, Vanhoucke V, Ioffe S, et al. [IEEE 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) - Las Vegas, NV, USA (2016.6.27-2016.6.30)] 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) - Rethinking the Inception Architecture for Computer Vision[J]. 2016:2818-2826.

[27] Johnson J, Alahi A, Fei-Fei L. Perceptual Losses for Real-Time Style Transfer and Super-Resolution[C].European Conference on Computer Vision. 2016

[28] Simonyan K, Zisserman A. Very Deep Convolutional Networks for Large-Scale Image Recognition[J]. Computer Science, 2014.

[29] Learned-Miller E, Huang G B, Roychowdhury A, et al. Labeled Faces in the Wild: A Survey[M]. Advances in Face Detection and Facial Image Analysis. Springer International Publishing, 2016.

[30] Liu Z, Luo P, Wang X, et al. Deep Learning Face Attributes in the Wild[C]. Proceedings of the IEEE international conference on computer vision. 2015: 3730–3738.

[31] Zhang K, Zhang Z, Li Z, et al. Joint Face Detection and Alignment Using Multitask Cascaded Convolutional Networks[J]. IEEE Signal Processing Letters, 2016, 23(10):1499-1503.

[32] Goodfellow I J. On distinguishability criteria for estimating generative models[J]. Statistics, 2014.

[33] Golub G H, Vorst H A V D. Eigenvalue computation in the 20th century[J]. Journal of Computational and Applied Mathematics, 2000, 123(1-2):35-65.

[34] Yoshida Y, Miyato T. Spectral Norm Regularization for Improving the Generalizability of Deep Learning[J]. arXiv preprint arXiv:1705.10941, 2017.

[35] Ma X, Zhang J, Qi C . Hallucinating face by position-patch[J]. Pattern Recognition, 2010, 43(6):2224-2236.

[36] Chen Z, Tong Y. Face Super-Resolution Through Wasserstein GANs[J]. arXiv preprint arXiv:1705.02438, 2017.