Pacific Graphics 2018
H. Fu, A. Ghosh, and J. Kopf
(Guest Editors)

Volume 37 (2018), Number 7

Generative Adversarial Image Super-Resolution Through Deep Dense Skip Connections

Xiaobin Zhu[1], Zhuangzi Li[1†], Xiaoyu Zhang[2†], Haisheng Li[1], Ziyu Xue[3] and Lei Wang[3]

[1]School of Computer and Information Engineering, Beijing Technology and Business University, Beijing, China
[2]Institute of Information Engineering, Chinese Academy of Sciences, Beijing, China
[3]Academy of Broadcasting Science, SAPPRFT, Beijing, China

Abstract

Recently, image super-resolution works based on Convolutional Neural Networks (CNNs) and Generative Adversarial Nets (GANs) have shown promising performance. However, these methods tend to generate blurry and over-smoothed super-resolved (SR) images, due to the incomplete loss function and powerless architectures of networks. In this paper, a novel generative adversarial image super-resolution through deep dense skip connections (GSR-DDNet), is proposed to solve the above-mentioned problems. It aims to take advantage of GAN's ability of modeling data distributions, so that GSR-DDNet can select informative feature representation and model the mapping across the low-quality and high-quality images in an adversarial way. The pipeline of the proposed method consists of three main components: 1) The generator of a novel dense skip connection network with the deep structure for learning robust mapping function is proposed to generate SR images from low-resolution images; 2) The feature extraction network based on VGG-19 is adopted to capture high frequency feature maps for content loss; and 3) The discriminator with Wasserstein distance is adopted to identify the overall style of SR and ground-truth images. Experiments conducted on four publicly available datasets demonstrate the superiority against the state-of-the-art methods.

CCS Concepts

•***Computing methodologies*** → ***Image processing;*** •***Computer systems organization*** → ***Neural networks;***

1. Introduction

Image super-resolution aims at reconstructing ground truth super-resolved (SR) images from low-resolution (LR) images, which has attracted increasing attention in research communities. Image super-resolution is often adopted for the reconstruction of license plate images [LLMC17], medical images [TLD*14], and any other conventional images [WYDCL18]. Nevertheless, this task is an inherently ill-posed problem for its one-to-many mapping nature. A variety of SR images can map to the same LR image, in contrast, the same LR images may have many different SR solutions. To deal with the problem, powerful and inferential models should be learned from LR to SR patches instead of traditional inflexible interpolation methods.

Conventional machine learning methods for single image super-resolution (SISR) can be broadly divided into two categories, namely internal example-based method and external example-based method. The internal example-based method [FF11, GBI09] directly exploited the self-similarity property and generated exemplar patches from the input image. And the external example-based method [KK10, YLC13] learned the mapping functions between LR and SR patches from external datasets. Recently, CNNs for image super-resolution was proposed in [DLHT16], which can potentially learned the mapping function and also built relevance with the previous external example-based method. VDSR successfully created a deeper CNN and realized a very fast convergence speed

Figure 1: *Comparing the state-of-the-art method (VDSR [KLL16a]) with the sharper, perceptually more plausible result produced by our method at 4× SR on an image from Set5.*

† Contact E-mail: lizhuangzii@163.com; zhangxiaoyu@iie.ac.cn.

using residual learning, which produced a very impressive promotion in SISR for its deep structure in [KLL16a]. Therefore, many researchers tried to build deep networks with various network structures like [KLL16b, TYL17]. In [TLLG17], the feature maps of each layer were propagated into all subsequent layers, providing an effective way to combine the low-level features and high-level features to boost the reconstruction performance, and alleviate the vanishing-gradient problem of very deep networks. However, these CNNs based methods are trained on the individual criteria of mean squared error (MSE), which only reflect the global mean loss and neglect the image details. Consequently, those methods are likely to generate blurry and over-smoothed SR images especially in $4\times$ upscaling factors as Figure 1 shows. In [LTH*17], SRGAN tried to solve the above-mentioned problem by using generative adversarial network. However, it failed to implement a powerful generative network and take the measurement of discriminator into consideration, which will resulting in collapsed or dim results.

In this paper, we propose a generative adversarial image super-resolution through deep dense skip connections (GSR-DDNet) to address above-mentioned problems and generate high-quality SR images from LR images. And for clarity, we use "SR" to denote the generated super-resolved images, "HR" to denote the ground-truth high resolution images, and "LR" to denote the input low resolution images. We make use of three different kinds of loss functions to measure SR and HR images in different levels, which can more highlight image details and consistent with human perception than previous individual criteria of MSE.

The main contributions can be concluded as follows:

- Firstly, a novel generative adversarial image super-resolution framework is proposed specially for learning mapping function from LR images to SR images. Specifically, the MSE loss, the content loss and the adversarial loss are combined and jointly optimized in an adversarial way.
- Secondly, a novel deep dense skip connection network for image super-resolution (SR-DDNet) is proposed as a powerful and robust generative model. The network adopts transition layers and small convolutional filters to save parameters for its deep structure. Meanwhile, dense connections and skip connections are used to solve vanishing-gradient problem to make the network easily to train.
- Last but not least, extensive experiments conducted on four benchmark datasets demonstrate the necessity of each step and the effectiveness of the proposed method. And the proposed method achieves superior performance against the state-of-the-art methods.

The rest of paper is organized as follows: in Sec. 2 we introduce some related works. The method is illustrated in Sec. 3,. In Sec. 4 the experimental results are shown and discussed. In Sec. 5, we draw a conclusion for the paper.

2. Related work

The related work is introduced from three aspects: (1) The conventional image super-resolution methods which utilize machine learning methods. (2) The CNN based image resolution method. (3) Image super-resolution utilizing generative adversarial networks.

2.1. Conventional methods for image super-resolution

Linear interpolation or bicubic interpolation methods were adopted for image super-resolution in the early years. Those methods tend to generate blurry super-resolved images for the lack of the prior knowledge. Recent years, The example-based methods were thoroughly researched, and those methods can be broadly divided into two categories, namely the internal example-based method and the external example-based method. The internal example-based method exploited the self-similarity property and generate exemplar patches from the input image [FF11, GBI09]. Cruz et al. proposed Wiener filter in similarity domain for super resolution, which formulated the SISR problem as a minimization of reconstruction error subject to a sparse self-similarity prior in [CMKE18]. On the other hand, the external example-based methods learned a mapping function from low resolution to high resolution patches from external datasets [KK10, YLC13, TSG13, YLC13]. The representative external example-based methods are sparse coding [YWHM08, BRGA12, HSA15, YWHM10, ZLW*14]. They payed particular attention to learning and optimizing the dictionaries or building efficient mapping functions from low-resolution to high resolution patches. Although, conventional methods can build models fast and easily, they cannot fit complex mapping functions to generate good results in the ill-posed single image super-resolution task.

2.2. Convolutional neural networks for image super-resolution

Recent years, CNNs have shown significant success in research communities [LBH15]. Some representative networks had exhibited their powerful capabilities in the large image classification task [KSH12, SZ14, SLJ*15, HZRS16, HLvdMW17]. Because of the great success in image classification, CNNs have also applied to single image super-resolution. In [DLHT16], SRCNN was proposed, and it can achieve better performance than those conventional methods utilizing machine learning techniques. Besides, it also demonstrated the relationship between sparse coding methods and itself. Afterwards, In [KLL16a], Kim et al. pursued deeper network named VDSR, and apparently gained better results in the single image super-resolution task. However, before the training step, the low-resolution (LR) image must be up-sampled to the same size as the labeled HR image. It would bring much computational load for the abundant convolutional operations on the large feature maps. Spontaneously, an effective sub-pixel convolution method was proposed in [SCH*16]. The inputs are LR pictures without interpolation operations, and the feature maps are up-sampled only in last serval layers. Some researchers also attempted to design novel network architectures for image super-resolution, e.g. the residual connection, the dense connection etc., and had shown better performance in [KLL16a, TYL17, LTH*17, TLLG17]. We also take advantage of this architecture to make our network get more powerful mapping capability. Most of above-mentioned methods trained their network using the mean squared error loss, which would lead to a high objective evaluation with over-smoothed image.

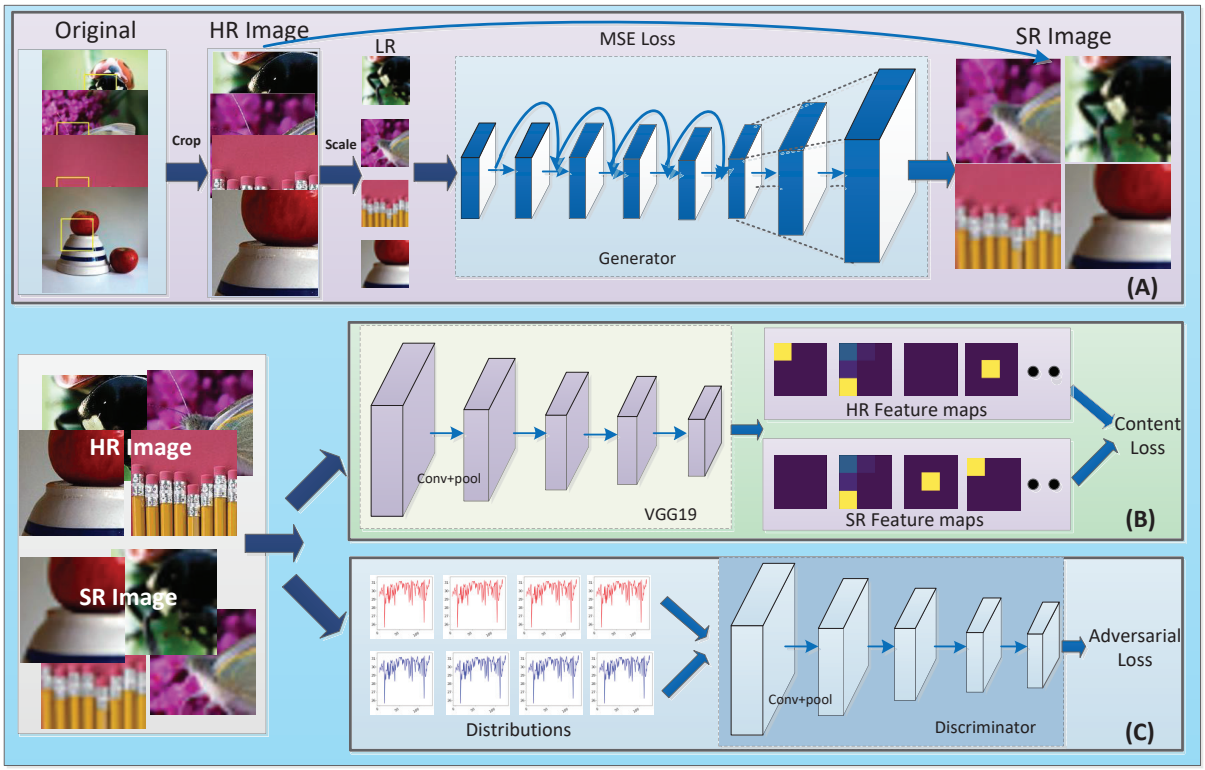

Figure 2: *The framework of the proposed method GSR-DDNet. (A): The generator named deep dense skip connection network (SR-DDNet). (B): The VGG-19 based feature extraction network. (C): The discriminator.*

2.3. Generative adversarial nets for image super-resolution

Pioneers utilized the auto-enconder [MMCS11] to generate images by a code. VAE [KW13] introduced the additional noise to make the auto-enconder more robust. Thanks to the recent development of Generative Adversarial Networks (GANs) [GPM*14], researchers have strived to automatically super-resolved image with GANs, which could be regarded as the game mechanism. Radford et al. introduced a class of CNNs called deep convolutional generative adversarial networks (DCGANs) in [RMC15]. Nowozin et al. proposed a unified framework, which revealed the essence of GANs and shown that any f-divergence can be used for training GANs in [NCT16]. In [ZML16], Zhao et al. proposed EBGAN which viewed the discriminator as an energy function that attributes low energies to the regions near the data manifold and higher energies to other regions. Mao et al. drew up a least-square error for the discriminator to solve the vanishing-gradient problem in [MLX*17]. In [ACB17], WGANs provided a more stable training framework. The Wasserstein distance or earth mover's distance rather than Jensen-Shannon divergence is concretely applied to the system, which is a better distance to measure two distributions similarity. However, the WGAN should clip weights for the discriminator, and it was likely to result a nonuniform distribution. Gulrajani et al. solved the problem by adding a penalty term in [GAA*17]. A unpaired image-to-image translation framework was proposed in [ZPIE17], which skillfully adopted GANs to implement image style transfer. In [WXH17], a comprehensive review on the GAN-based methods was provided. GANs were also introduced in image super-resolution for generating more realistic super-resolved images from low-resolution images in [LTH*17, SSH17]. But there is lack of consideration on the architecture of the generator and the distance measured by the discriminator, which is like to generate dim or collapsed generated images. Our method overcomes these disadvantages. Advanced generative network and stable distance are both adopted for generating high-quality super-resolved images.

3. The proposed method

Figure 2 shows the framework of the proposed method, the generator, i.e., the deep dense skip connection network (SR-DDNet) aims to generate super-resolved (SR) images, and the discriminator aims to discriminate SR images are fake, the VGG-19 based feature extraction network is used for comparing the content of features. The framework is jointly trained, and all losses are used to update parameters of generator.

The framework mainly contains three components. In the Figure 2 (A), original input images are randomly cropped by a settled size, and we resize them to LR images. Then, LR images are put into the generator for achieving SR images. Comparing SR images with the corresponding ground-truth images, we get the mean squared error (MSE) loss. The Figure 2 (B) adopts the VGG-19

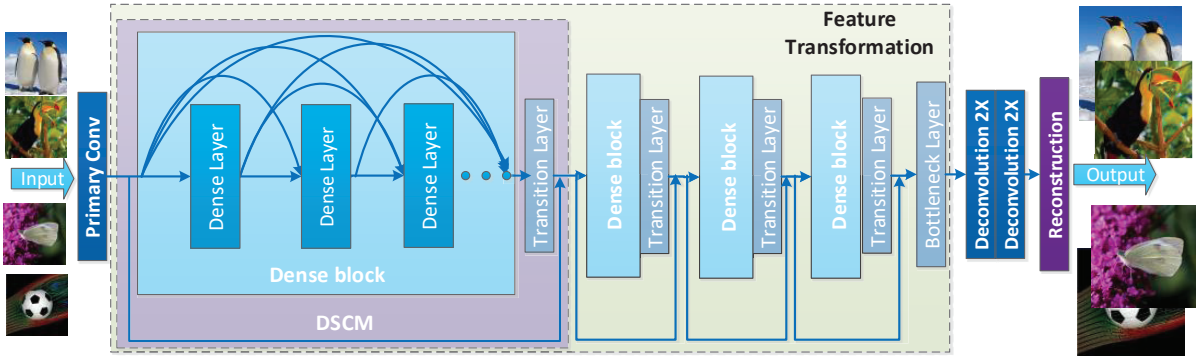

Figure 3: *The architecture of SR-DDNet. The framework contains 4 parts, the primary convolutional layer, the feature transformation, the deconvolution layer and the reconstruction layer.*

as the feature extraction network without loss of generality. The content loss is computed by the MSE between the feature maps of SR and HR. And in the Figure 2 (C), the Wasserstein distance is adopted to evaluate the difference of RGB space between the SR images and HR counterparts. After convolutional and pooling operations, the discriminator can output the distance between two distributions for computing the adversarial loss. The three parts have three different loss functions, when training the generator, the feature extraction network and discriminator are fixed, they only provides the loss value. The generative network are optimized by back propagation through the three combined loss functions.

In Sec. 3.1, the generative network called deep dense skip connection network for image super-resolution (SR-DDNet) is introduced, and we also analyze its internal architecture and relational calculating operations. We detailedly describe loss functions in Sec. 3.2. And in Sec. 3.3, the detailed training process is explained.

3.1. Generator

Our generator aims to recover high-resolution (HR) images from low-resolution (LR) images. Given training low-resolution LR images I^{LR} and corresponding ground-truth HR images I^{HR}, we define a sophisticated nonlinear mapping function $F : I^{LR} \rightarrow I^{HR}$. The destination of the method is to optimize the F by learning pairs of I^{LR} and I^{HR}. In fact, the F can be approached by the generator G_θ, where the θ is trainable parameters of the generative network, and the goal of single image super-resolution can be described by the following formula:

$$\hat{\theta} = \arg\min_{\theta} \; L(\mathrm{G}_\theta(I^{LR}) \, , \, I^{HR}). \tag{1}$$

L is a loss function computing the error of super-resolved (SR) images and ground truth HR images, and the purpose of SR is to optimize θ for a minimize loss. In our framework, G represents the proposed deep dense skip connection network (SR-DDNet), and loss L formulated as a kind of hybrid loss.

Network architecture The architecture of the CNN plays an important role in single image super-resolution. Inspired by SR-DenseNets [TLLG17], we propose a novel architecture employing deep dense skip connections called SR-DDNet. As shown in Figure 3, numerous dense connections and skipped connections are adopted and designed. SR-DDNet mainly contains 4 components. In the first part, the primary convolution, who adopts double 3×3 convolutional layers, aims to extract the edges and primary features and provides a fundamental local receptive field. The second part is the feature transformation, which are composed of many dense layers for sufficient learning efficiency. Primary features can be transformed to another space that they are easily to reconstruct HR images. Each dense block and transition layer are combined a module, we call it dense skip connection module (DSCM). At the third part, double deconvolution $2\times$ layers are used to magnify feature maps to make sure the feature maps have the same size with HR images. The last part is the reconstruction layer taking 3×3 convolutional filters, can recover the input channels (RGB 3 channels) of images.

Dense skip connection module Dense skip connection module (DSCM) is the main component of SR-DDNet. Fig .4 shows the architecture of the DSCM block, Batch Normalizations (BNs) and Rectified Linear Units (ReLUs) are used. Two convolutional layers are cascaded in each dense layer, the first 3×3 convolutional kernel is used for large receptive field. The second 1×1 convolution is adopted to deepen the network so that the network can learn

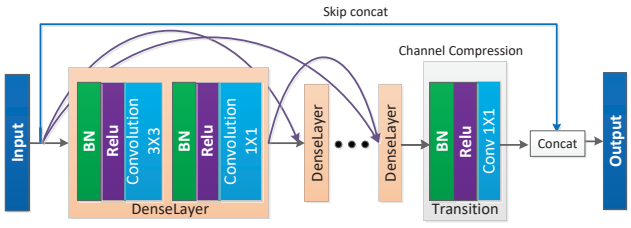

Figure 4: *The architecture of a dense skip connection module.*

a robust feature representation, and the small filter will not produce too many parameters. Dense layers are sequentially stacked, they have short paths from previous dense layers. Consequently, the $\ell-th$ dense layer receives the feature-maps of all preceding layers, $x_0,...,x_{\ell-1}$ as the input:

$$x_\ell = H_\ell([x_0,x_1,...,x_{\ell-1}]), \tag{2}$$

where $[x_0,x_1,...,x_{\ell-1}]$ is feature maps concatenation for layers $0,...,\ell-1$. It can be seen that with more operations of dense connections, there are more feature maps will be concatenated that will bring more parameters and computation loads for the following layers. To solve this problem, a transition layer is adopted. The transition layer is composed of a BN layer, an activation layer of ReLU and a convolutional layer with 1×1 filters. The aim of the transition layer is not only for the nonlinear mapping but also for compressing feature maps, which can output a small number of feature maps. DSCM adopts the skipped connection for combining input and output feature maps, which guarantees transmission from the low-level feature to high-level semantic feature after many convolution operations. The operation can be represented by:

$$x_{output} = [x_0, T(x_l)], \tag{3}$$

where the x_0 is the input from the first dense layer, x_ℓ is the output from the last dense layer. $T(x)$ is the function realized by the transition layer. The output of DSCM inferences to next module or the bottleneck layer.

3.2. Loss functions

The training loss function is a hybrid loss which consists of a mean squared error (MSE) loss, a content loss, and an adversarial loss. We call it ℓ_{total}, it can be written as the following formula:

$$\ell_{total} = l_{mse} + \lambda_1 \cdot l_{content} + \lambda_2 \cdot l_{adv}, \tag{4}$$

where λ_1 and λ_1 is hyper-parameters used for balancing weights of each loss. Each loss has their own unique effect. The MSE loss is basal loss and it concentrates on mean pixel-wise errors. The content loss is used to guarantee two image have the same high-frequency features. The adversarial loss can discriminate whether two images have the same style.

Mean squared error loss and content loss The pixel-wise mean squared error (MSE) loss is a most conventional loss function in super-resolution task, which aims to compare the difference on the pixel-level. Early works usually adopt the single MSE loss. The loss is significant in our method, which is used to lead the main training direction, and also straightway influence the pixel-wise effect of our SR images. For the $i-th$ low-resolution (LR) image I_i^{LR} and its ground truth HR image I_i^{HR}, their MSE loss can be written as:

$$\ell_{mse} = \frac{1}{WHC}||G(I_i^{LR}) - I_i^{HR}||_2^2, \tag{5}$$

where W and H are width and height of I_i^{HR}, C is the channel count of I_i^{HR}. This loss directly measures the gap between SR images and HR images, which is helpful for leading the macroscopical integral training direction.

CNNs can get rich various semantic information after several convolutional layers. Thus, in this design, a pre-trained VGG-19 network is viewed as the feature extractor, and we only use the high-level feature maps where they come from the last max pooling layer. Besides, VGG-19 are untrainable and it only works as a invariant mapping function from images to content features. These features can reveal semantic information clearly, which provide high-frequency comparisons to produce a sharp result. I^{SR} is SR the image generated by the densely network. For the $i-th$ LR and HR, The content loss can be described as:

$$\ell_{content} = \frac{1}{whc}||\phi(I_i^{SR}) - \phi(I_i^{HR})||_2^2. \tag{6}$$

ϕ is the feature extraction function. w and h are the width and height of the I_i^{HR}, the c is the channel count of the I_i^{HR}.

Adversarial loss GANs provide a adversarial training way to improve the generation quality from random noises. The similarity is measured by the discriminator between the generated distribution and the target distribution. In the GANs system, given a generated image, the discriminator should identify as fake. On the other hand, the generator aims to generate images which make the discriminator feel they are real images. In the super-resolution task, the input to the generator is a low-resolution image instead of the random noise. The adversarial learning method is used to guarantee real images can be generated rather than over-smoothed or over-sharped images. Thus, discriminator plays an essential role in our GANs system, which measures similarities between generated SR images and ground truth HR images. Previous GANs adopt JS-divergence to discriminate the target distribution and generated distribution. We assume the real data distribution as $\mathbb{P}_r$, and the generated distribution $\mathbb{P}_g$. The JS-divergence can be described as the following formula:

$$JS(\mathbb{P}_r,\mathbb{P}_g) = KL(\mathbb{P}_r \parallel \mathbb{P}_m) + KL(\mathbb{P}_g \parallel \mathbb{P}_m). \tag{7}$$

$\mathbb{P}_m$ is equal to $(\mathbb{P}_r+\mathbb{P}_g)/2$. The KL-divergence can be written by the following formula:

$$KL(\mathbb{P}_r,\mathbb{P}_g) = \int \log(\frac{P_r(x)}{P_g(x)})P_r(x)dx. \tag{8}$$

According to theories proposed by Goodfellow et al. [GPM*14], the JS-divergence based discriminator should optimize the following problem in the image super-resolution task:

$$\max_D \ \mathbb{E}_{I\sim p_{HR}}[\log D(I)] + \mathbb{E}_{I\sim p_{SR}}[\log(1-D(I))], \tag{9}$$

where the $I\sim p_{HR}$ denotes images sampled from the HR distributions. And $I\sim p_{HR}$ denotes images sampled from super-resolved (generative) image distributions. And the Generator aims to optimize the following formula:

$$\min_G \ \mathbb{E}_{I\sim p_{LR}}[\log(1-D(G(I)))]. \tag{10}$$

$I\sim p_{LR}$ means that we sample images from low-resolution image distributions.

However, JS-divergence is bounded between 0 and $\log 2$, and cannot measure two totally different distributions because their divergence is beyond the boundary. Therefore, training GANs is especially sensitive to hyper-parameters and requires carefully design networks. Sometimes it would likely to result in mode collapse.

Table 1: *The comparisons with the state-of-the-art methods by PSNR and SSIM (4×). Scores in bold denote the highest values.*

Dataset	Bicubic	A+ [TSG14]	SCN [WLY*15]	SelfExSR [HSA15]	SRCNN [DLHT16]	VDSR [KLL16a]	DRCN [KLL16b]
Set5	28.42 / 0.810	30.30 / 0.859	30.39 / 0.862	30.33 / 0.861	30.49 / 0.862	31.35 / 0.882	31.53 / 0.884
Set14	26.10 / 0.704	27.43 / 0.752	27.48 / 0.751	27.54 / 0.756	27.61 / 0.754	28.03 / 0.770	28.04 / 0.770
BSD100	25.96 / 0.669	26.82 / 0.710	26.87 / 0.710	26.84 / 0.712	26.91 / 0.712	27.29 / 0.726	27.24 / 0.724
Urban100	23.15 / 0.659	24.34 / 0.720	24.52 / 0.725	24.82 / 0.740	24.53 / 0.724	25.18 / 0.753	25.14 / 0.752
Average	25.90 / 0.710	27.22 / 0.760	27.32 / 0.762	27.38 / 0.767	27.39 / 0.763	27.96 / 0.782	27.99 / 0.782

Dataset	DRRN [TYL17]	LapSRN [LHAY17]	MemNet [TYLX17]	SRDenseNet [TLLG17]	SR-DDNet (ours)	GSR-DDNet- (ours)	GSR-DDNet (ours)
Set5	31.68 / 0.888	31.54 / 0.885	31.74 / 0.889	32.02 / 0.893	**32.21 / 0.903**	31.34 / 0.885	31.99 / 0.897
Set14	28.21 / 0.772	28.19 / 0.772	28.26 / 0.772	28.50 / 0.778	**28.71 / 0.789**	28.10 / 0.770	28.51 / 0.780
BSD100	27.38 / 0.728	27.32 / 0.728	27.40 / 0.728	27.53 / 0.733	**27.69 / 0.743**	27.23 / 0.728	27.60 / 0.736
Urban100	25.44 / 0.763	25.21 / 0.756	25.50 / 0.763	26.05 / 0.782	**26.18 / 0.790**	25.29 / 0.785	26.01 / 0.780
Average	28.18 / 0.788	28.07 / 0.786	28.23 / 0.789	28.52 / 0.796	**28.70 / 0.806**	27.99 / 0.792	28.53 / 0.798

Wasserstein GANs [ACB17] is an alternative method for GANs. It can improve the stability of learning, get rid of problems like mode collapse, and provide meaningful learning curves for debugging and hyper-parameter searching. The Wasserstein distance can be written as:

$$W(\mathbb{P}_r,\mathbb{P}_g) = \inf_{\gamma\in\prod(\mathbb{P}_r,\mathbb{P}_g)} \mathbb{E}_{x,y\sim\gamma}[c(x,y)], \tag{11}$$

where $\prod(\mathbb{P}_r,\mathbb{P}_g)$ is the set of all joint distributions $\gamma(x,y)$ whose marginals respectively are $\mathbb{P}_r$ and $\mathbb{P}_g$. Intuitively, $\gamma(x,y)$ indicates how much "mass" must be transported from x to y in order to transform the distributions $\mathbb{P}_r$ into the distribution $\mathbb{P}_g$. According to [ACB17], this distance can be applied into the GANs system, for the discriminator D, it can be converted to the following optimization:

$$\max_{||D||_L\leq 1} \mathbb{E}_{I\sim p_{HR}}[D(I)] - \mathbb{E}_{I\sim p_{SR}}[D(I)]. \tag{12}$$

$||D||_L \leq 1$ limits the discriminator must be a 1- Lipschitz function. Generally speaking, it restricts the discriminator cannot output overlarge values. However, the 1- Lipschitz function cannot effectively be represented in the training procedure. In the early experiments, the weight clipping was adopted to approximate the 1- Lipschitz constraint, but the method usually result in the K- Lipschitz function rather than the 1- Lipschitz, it also brings inferior parameters distribution. WGAN-GP [GAA*17] efficiently resolved the problem using regular terms. we reference it for our training. when we train a generator, the parameters of the discriminator should be fixed, and we define the adversarial loss for the generator:

$$\ell_{adv} = -\mathbb{E}_{I\sim p_{LR}} D(G(I)), \tag{13}$$

where the I denotes images from the low-resolution distribution, and the generator should make l_{adv} as small as possible.

3.3. Training process

The proposed method is trained via an adversarial process, we firstly get training patches from the training dataset, the low-resolution (LR) patch should be put into the generator G to generate super-resolved (SR) image patches, and we fix the generator G to train the discriminator D, the D discriminates SR images and high-resolution (HR) images according to formula (12) and update the parameters of D by Adam. Then, we calculate the hybrid loss for training the generator G. The MSE loss l_{mse} can be directly computed by SR and HR patches, we put these patches into the fixed VGG-19, and get feature maps of the SR and HR patches to compute the content loss $l_{content}$. As for the adversarial loss, we fix the discriminator D, put HR and SR patches into it, and get the adversarial loss l_{adv}. l_{mse}, $l_{content}$ and l_{adv} are combined to l_{total}, The G needs to minimize l_{total} function as formula (4) to update its parameters by Adam. We do above operations many epochs, and we will get a convergent model. The detailed algorithm is shown in Algorithm 1.

Algorithm 1 Training on **GSR-DDNet**. the clip size c is hyper-parameters, and is set as 120. We set $\alpha = 1.0$ $\beta = 0.01$ $\gamma = 0.001$

1: **for** number of training epochs **do**
2: **for** k steps **do**
3: Sample batch of HR images $I_1^{HR},...,I_m^{HR}$ from the HR dataset.
4: Randomly crop $c \times c$ I^{HR} and zoom them to 1/4 so that get batch of LR images samples $I_1^{LR},...,I_m^{LR}$.
5: Update the discriminator by Adam.
$\nabla_{\theta_d} \frac{1}{m}\sum_{i=1}^{m}[D_\theta(I_i^{HR})] - [D_\theta(G(I_i^{LR}))] + \lambda(||\nabla_I D_\theta(I)||_2)^2$
6: Reuse the batch of HR sample and LR sample $I_1^{LR},...,I_m^{LR}$ from step 4. We calculate three losses ℓ_{MSE}, $\ell_{content}$ and ℓ_{adv}:
$\ell_{MSE} = \frac{1}{m}\sum_{i=1}^{m}\frac{1}{WHC}(G(I_i^{LR}) - I_i^{HR})^2$,
$\ell_{content} = \frac{1}{m}\sum_{i=1}^{m}\frac{1}{whc}(\phi(I_i^{SR}) - \phi(I_i^{HR}))^2$
$\ell_{adv} = -\frac{1}{m}\sum_{i=1}^{m} D(G(I_i^{LR}))$
7: Fix the D and update the generator G by Adam.
$\nabla_{\theta_g} \alpha \cdot l_{MSE} + \beta \cdot l_{content} + \gamma \cdot l_{adv}$
8: **end for**
9: **end for**

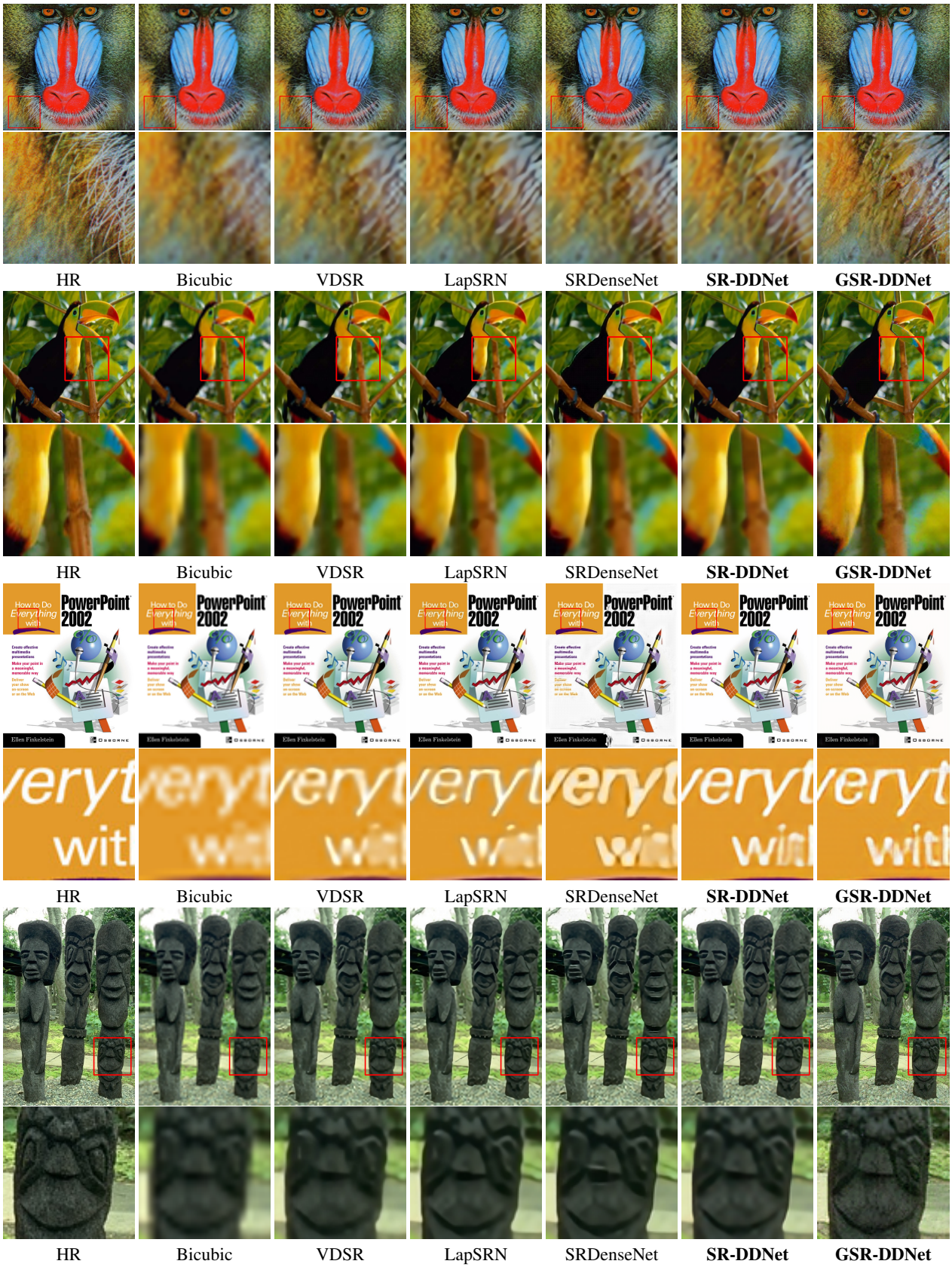

Figure 5: *The visual comparisons on the Set5, Set14, and BSD100 with scaling factors as 4.*

4. Experiments

In our algorithm, 60,000 images are randomly selected from ImageNet to build the training dataset. To validate the effectiveness and robustness of the proposed algorithm, we conduct experimental test on four publicly available datasets: Set5 [BRGA12], Set14 [ZEP10], BSD100 [MFTM01] and Urban100 [HSA15]. Details are shown in the following subsections.

4.1. Dataset and metrics

The dataset Set5 [BRGA12] and Set14 [ZEP10], which contain 5 and 14 different types of images. The Berkeley segmentation dataset [MFTM01], namely BSD100, includes 100 natural images for testing. And the Urban100 contains 100 images of the urban scenario [HSA15]. All experiments are performed using 4× upscaling factors from low resolution to high resolution. The peak signal-to-noise ratio (PSNR) and the structural similarity (SSIM) index are two criterion metrics for objective evaluation. According to the former works, the PSNR and SSIM are calculated on the Y-channel of images. The mean opinion score (MOS), from uninfluenced 15 raters, is adopted as the criteria of subjective evaluation, which range from 1 (lowest quality) to 5 (best quality).

4.2. Compare with other approaches

To exhibit the experiment effect, we take the proposed method to compare with state-of-the-art methods evaluated on PSNR and SSIM. Notably, the proposed method SR-DDNet represents generator only trained on MSE loss. To further demonstrate the superiority of the proposed method with Wasserstein distance, "GSR-DDNet-" is proposed as a comparison, in which JS divergence instead of Wasserstein distance is implemented in GAN's training, and the GSR-DDNet adopts WGAN-GP based training. Some of classical methods are selected for comparison like: Bicubic, A+ [TSG14], SCN [WLY*15], SelfExSR [HSA15], SRCNN [DLHT16], VDSR [KLL16a], DRCN [KLL16b], DRRN [TYL17], LapSRN [LHAY17], MemNet [TYLX17] and SRDenseNet [TLLG17]. As shown in Table 1, SR-DDNet shows highest value that it achieves 32.21 / 0.903 scores (on Set5), while SRCNN obtains 30.49 / 0.862, and SR-DDNet outperforms them about 1.72 and 0.041 scores. Comparing with SRDenseNet, SR-DDNet promotes it about 0.2 / 0.01. On Set14 dataset, SR-DDNet outperforms DRCN and DRRN about 0.67 / 0.022 and 0.5 / 0.017. SR-DDNet gets 27.69 / 0.743 scores (on BSD100), while VDSR only gets 27.29 / 0.726. On Urban100 dataset, it can be seen that SR-DDNet arrives 26.18 / 0.790 scores, and SRDenseNet makes a very close value to it. Compared with scores of SR-DDNet, PSNR/SSIM of GSR-DDNet fall by 0.2 / 0.006 (on Set5), 0.2 / 0.01 (on Set14), 0.1 / 0.007 (on BSD100) and 0.2 / 0.1 (on Urban100) due to the content and adversarial losses which are irrelevant to MSE. While GSR-DDNet- decreases more. However, objective comparisons are not conclusive for the experiment, we should further evaluate the visual result of generated image.

The MOS is adopted for subjective comparison. we select SRDenseNet [TLLG17], SR-DDNet, GSR-DDNet- and GSR-DDNet As shown in Table. 2, GSR-DDNet makes highest scores on Set5 and outperforms other methods about 0.3, 0.2, and 0.15. GSR-DDNet obtains 3.34 (on Set14), while SRDenseNet only gets 2.55. There are more promotions on BSD100 dataset, the proposed method GSR-DDNet- and GSR-DDNet are 3.22 and 3.89 respectively, which obviously exceed other methods. On Urban100 dataset, GSR-DDNet also achieves highest scores 3.65, while other methods are all lower than 3.1. Comparing with total average scores, GSR-DDNet outperforms other methods with significant scale of 0.86, 0.7 and 0.46 respectively. Thus, GSR-DDNet makes the highest subjective evaluation, which means that it generates the best visual quality images. From those subjective and objective comparisons, we conclude that the adversarial learning procedures can improve the visual quality of the super-resolved images while constrain the PSNR/SSIM scores. Besides, the proposed GSR-DDNet is superior to GSR-DDNet- (GAN based), which obtains best visual result with only a little decreases of PSNR. Here, some representative methods like Bicubic, VDSR [KLL16a], LapSRN [LHAY17], SRDenseNet [TLLG17], SR-DDNet and GSR-DDNet are selected to exhibit their effect in local detail textures, as shown in Figure 5, Those pictures are the "baboon" from Set14, "bird" from Set5, "ppt" from Set14, "image001" from BSD100. Figure 6 and Figure 7 show another visual results, in which whole images are shown instead of local details.

Table 2: *The evaluation on the Set5, Set14, BSD100 and Urban100 datasets, with the up-scaling factor set as 4.*

Dataset	SRDenseNet [TLLG17]	SR-DDNet (ours)	GSR-DDNet- (ours)	GSR-DDNet (ours)
Set5	2.71	2.79	2.85	**3.02**
Set14	2.55	2.64	2.94	**3.34**
BSD100	2.56	2.98	3.22	**3.89**
Urban100	2.67	2.69	3.08	**3.65**
Average	2.62	2.78	3.02	**3.48**

4.3. Implementation details and timing

The proposed network is trained specially for 4-scale factor super-resolution. Randomly crop and randomly flip are set for the training dataset. The crop size is 120×120 for each image, thus the input LR patches' size is 30×30. For only one channel image, we convert it to RGB channel as the input. The primary convolution adopts double 3×3 convolution with 32 and 64 channels of outputs. In the feature transformation part, 4 dense skip connection modules (DSCMs) are used, in which compression rate is set to 0.5. The 4 DSCMs have different numbers of dense layers, which are 6, 12, 24 and 16, respectively. Growth rate is set to 32 for dense connection. Each dense layer includes a 3×3 and a 1×1 convolutional layers. Thus, the network has 123 convolutional layers in total. The ReLU is used as the activation function expect except the up-sample layer, which adopts PReLU. For the bottle neck layer, it outputs 256 feature maps with 1×1 filter size. The input of the reconstruction layer contains 256 feature maps and outputs 3. The discriminator is built with 8 convolutional layers, and unites BN and LeakyReLU with negative slope 0.2. For optimization, the proposed method is optimized by Adam [KB14] with learning rate 0.0001. For each 5 iterations, the learning rate will increase by the

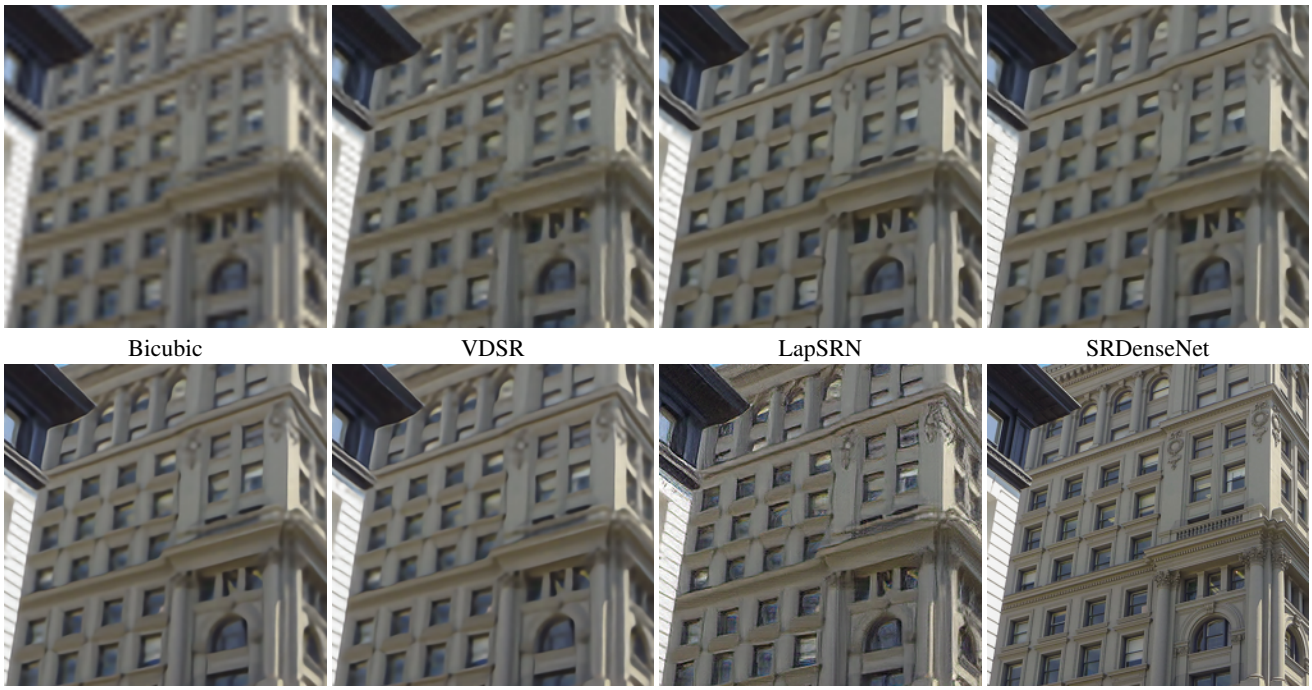

Figure 6: *The visual comparisons on Urban100 "img014" with scaling factors as* $4\times$.

scale of 0.9. Our training process stopped when the training reaches 200 epochs. Experiments are performed on double NVIDIA Titan X GPUs for training and testing.

It takes about 36 hours to train the SR-DDNet (MSE loss only). As for the GSR-DDNet, it takes about 50 hours. They both have identical testing time about 0.16s per image on Set5 dataset. The theoretical time complexity is $O(\sum_{l=1}^{D} M_l K_l C_{l-1} C_l)$, where D is depth of our framework, l is the l-th layer, M is length of feature map and K is length of kernel. C represents the convolution kernel number.

4.4. Discussion and future work

Experiment discussion We discuss experiment results in three aspects. (1) The highest objective evaluation (PSNR / SSIM) of SR-DDNet. Objective evaluation is built on mean squared error (MSE). Thus, when a network can learn a small MSE, they can get a high objective evaluation like PSNR. Therefore, the learning ability of the network is essential. We know that when substantially increasing the depth of networks, the more powerful learning ability would we have. Ordinary CNNs based methods cannot stack too many layers to depth for hard training. In here, our network adopts dense connections and skip connections, which bring more deep structure and do not necessarily think about the vanishing gradient problem. And vast training data is suit for our deep structure rather than other shallow of networks. What's more, our networks adopt many 1×1 convolutional layer that increased layers of network without generating too many parameters. (2) High PSNR and high-quality images. The MSE is a mean error of whole image, it cannot only describe the truth result of one image. The difference form a point-to-point pixel can be neutralized by other surrounding pixels. In another point, our clear and acceptable generated image is only a small parts of MSE solution space that is scarcely possible to find it for its enormous solution space. When we add the content loss and the adversarial loss to train generator, it is equal to two regularization items, which make the optimizer uncomfortably find a best MSE solution. We can see the GAN based training procedure is not stable than only MSE training process. However, the GAN based methods allow the model to have more "imagination" for latent semantic information. Thus, it can take some results more coincide with human vision. (3) Wasserstein distance to discriminator. Wasserstein distance is a improvement of JS-divergence based GANs. it could produce concrete difference value instead of bounded JS-divergence, which really estimates how far or how close about two images distributions. Generally speaking, Wasserstein distance stably and effectively provides specific direction for update parameters and keeps a balance training of generator and discriminator.

Difference between SR-DDNet and SRDenseNet Our SR-DDNet and SRDenseNet [HLvdMW17] both adopt densely connected in network structure. However, three main differences between SRDenseNets and our SR-DDNets. (1) DSCMs introduce a compressed layer that compress numbers of output channels from the dense block output in proportion. But SRDenseNet use settled numbers of output in each dense block. In other words, SR-DDNets can better retain high-level feature maps, which will gain more semantic information. (2) There are two different types convolutional

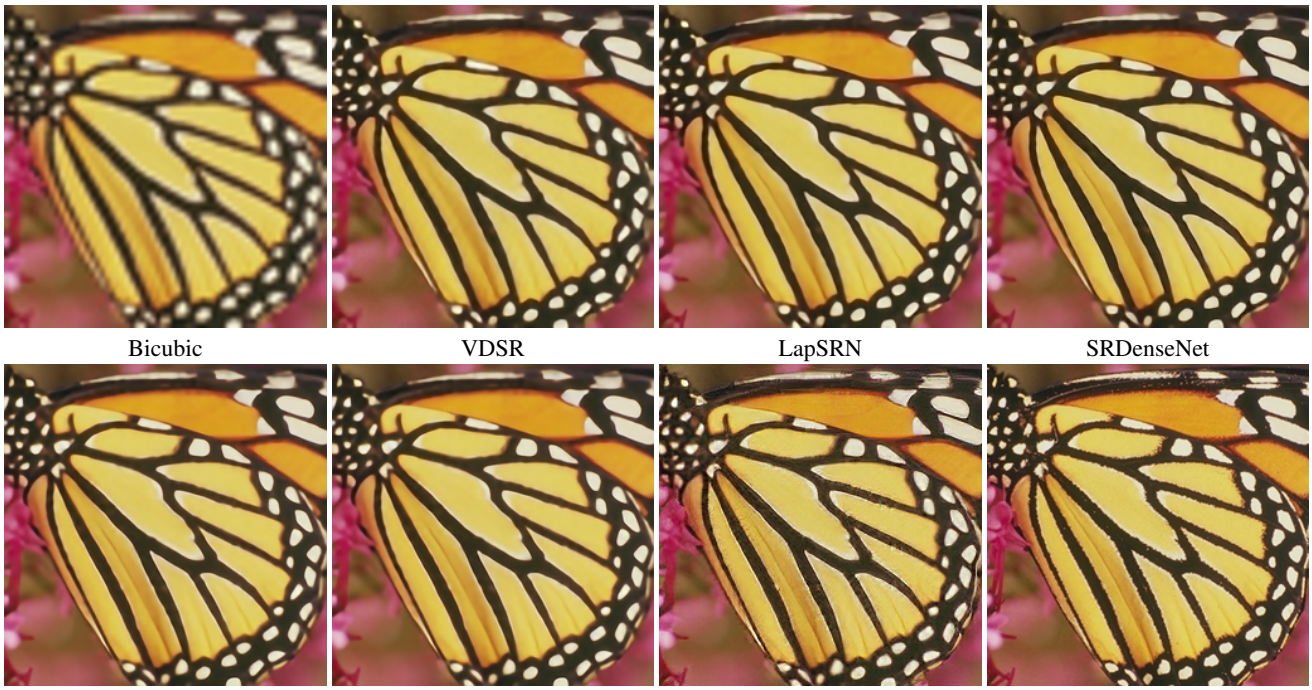

Figure 7: *The visual comparisons on Set5 "butterfly" with scaling factors as* $4\times$.

layers are adopted in each dense layer. For SRDenseNet, only a 3×3 convolutional layer is used in a dense block, which reduce the capacity of data fitting. (3) We propose a general dense skipped connection structure, and the proposed SR-DDNet achieves 123 convolutional layers while SRDenseNet only has 65 in total. In conclusion, SR-DDNet is a improvement of SRDenseNet, which can better learning a LR to HR transformation.

Difference between GSR-DDNet and SRGAN GSR-DDNet and SRGAN [LTH*17] both adopt GANs architecture, but there are two mainly different between our framework and SRGAN, The generator structure is different, SRGAN adopts ResNets structure that residual learning is used, but we apply dense connection and skip connection to our network. The next difference is discriminator, the JS-divergence is used to discriminate the gap between two images. However, we adopt a more robust and concrete Wasserstein distance to measure it.

Future works In our future works, we intend to research the improvement of the proposed method in high PSNR and high-quality SR images simultaneously, we plan to solve deal the problem by design a novel local loss function to make SR image more specific in image local details. Although we adopt some tracks to reduce parameters, the generative network still has some redundancies, we intend to simplify its structure by adopting recursive architecture. The next research direction is that we apply super-resolution method as a fixed generative model to high-resolution image processing as [WLX17, ZXZ*16, Zha16, LWD*17, LLLL15].

5. Conlusion

In this paper, we propose a generative adversarial image super-resolution method through deep dense skip connections. In the framework, a novel deep dense skip connection network, served as the generator, is utilized to generate SR images. The network has powerful and robust fitting ability for it makes full use of hierarchical features to deepen the structure. And a discriminative network with Wasserstein distance is used to calculate the distance between SR images and HR images. Three loss functions, including a mean squared error loss of pixels, a content loss calculated by a VGG19 network and an adversarial loss calculated by the discriminator, are unified together for training high performance generator. Experimental comparisons on four public available benchmark datasets show that the proposed framework GSR-DDNet outperforms the state-of-the-art algorithms and generates high-quality SR images.

Acknowledgment

This work was supported by National Key R&D Program of China (2017YFB1401000) and National Natural Science Foundation of China (61501457, 61602517, 61877002). The corresponding authors are Zhuangzi Li and Xiaoyu Zhang, who contribute equally to this paper.

References

[ACB17] ARJOVSKY M., CHINTALA S., BOTTOU L.: Wasserstein generative adversarial networks. In *Proceedings of the 34th International Conference on Machine Learning, ICML 2017, Sydney, NSW, Australia, 6-11 August 2017* (2017), pp. 214–223. 3, 6

[BRGA12] BEVILACQUA M., ROUMY A., GUILLEMOT C., ALBERI-MOREL M.: Low-complexity single-image super-resolution based on nonnegative neighbor embedding. In *British Machine Vision Conference, BMVC 2012, Surrey, UK, September 3-7, 2012* (2012), pp. 1–10. 2, 8

[CMKE18] CRUZ C., MEHTA R., KATKOVNIK V., EGIAZARIAN K. O.: Single image super-resolution based on wiener filter in similarity domain. *IEEE Trans. Image Processing 27*, 3 (2018), 1376–1389. 2

[DLHT16] DONG C., LOY C. C., HE K., TANG X.: Image super-resolution using deep convolutional networks. *IEEE Trans. Pattern Anal. Mach. Intell. 38*, 2 (2016), 295–307. 1, 2, 6, 8

[FF11] FREEDMAN G., FATTAL R.: Image and video upscaling from local self-examples. *ACM Trans. Graph. 30*, 2 (2011), 12:1–12:11. 1, 2

[GAA*17] GULRAJANI I., AHMED F., ARJOVSKY M., DUMOULIN V., COURVILLE A. C.: Improved training of wasserstein gans. In *Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, 4-9 December 2017, Long Beach, CA, USA* (2017), pp. 5769–5779. 3, 6

[GBI09] GLASNER D., BAGON S., IRANI M.: Super-resolution from a single image. In *IEEE 12th International Conference on Computer Vision, ICCV 2009, Kyoto, Japan, September 27 - October 4, 2009* (2009), pp. 349–356. 1, 2

[GPM*14] GOODFELLOW I. J., POUGET-ABADIE J., MIRZA M., XU B., WARDE-FARLEY D., OZAIR S., COURVILLE A. C., BENGIO Y.: Generative adversarial nets. In *Advances in Neural Information Processing Systems 27: Annual Conference on Neural Information Processing Systems 2014, December 8-13 2014, Montreal, Quebec, Canada* (2014), pp. 2672–2680. 3, 5

[HLvdMW17] HUANG G., LIU Z., VAN DER MAATEN L., WEINBERGER K. Q.: Densely connected convolutional networks. In *2017 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, Honolulu, HI, USA, July 21-26, 2017* (2017), pp. 2261–2269. 2, 9

[HSA15] HUANG J., SINGH A., AHUJA N.: Single image super-resolution from transformed self-exemplars. In *IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2015, Boston, MA, USA, June 7-12, 2015* (2015), pp. 5197–5206. 2, 6, 8

[HZRS16] HE K., ZHANG X., REN S., SUN J.: Deep residual learning for image recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2016, Las Vegas, NV, USA, June 27-30, 2016* (2016), pp. 770–778. 2

[KB14] KINGMA D. P., BA J.: Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980* (2014). 8

[KK10] KIM K. I., KWON Y.: Single-image super-resolution using sparse regression and natural image prior. *IEEE Trans. Pattern Anal. Mach. Intell. 32*, 6 (2010), 1127–1133. 1, 2

[KLL16a] KIM J., LEE J. K., LEE K. M.: Accurate image super-resolution using very deep convolutional networks. In *2016 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2016, Las Vegas, NV, USA, June 27-30, 2016* (2016), pp. 1646–1654. 1, 2, 6, 8

[KLL16b] KIM J., LEE J. K., LEE K. M.: Deeply-recursive convolutional network for image super-resolution. In *2016 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2016, Las Vegas, NV, USA, June 27-30, 2016* (2016), pp. 1637–1645. 2, 6, 8

[KSH12] KRIZHEVSKY A., SUTSKEVER I., HINTON G. E.: Imagenet classification with deep convolutional neural networks. In *Advances in Neural Information Processing Systems 25: 26th Annual Conference on Neural Information Processing Systems 2012. Proceedings of a meeting held December 3-6, 2012, Lake Tahoe, Nevada, United States.* (2012), pp. 1106–1114. 2

[KW13] KINGMA D. P., WELLING M.: Auto-encoding variational bayes. *CoRR abs/1312.6114* (2013). arXiv:1312.6114. 3

[LBH15] LECUN Y., BENGIO Y., HINTON G. E.: Deep learning. *Nature 521*, 7553 (2015), 436–444. 2

[LHAY17] LAI W., HUANG J., AHUJA N., YANG M.: Deep laplacian pyramid networks for fast and accurate super-resolution. In *2017 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, Honolulu, HI, USA, July 21-26, 2017* (2017), pp. 5835–5843. 6, 8

[LLLL15] LI C., LIU Q., LIU J., LU H.: Ordinal distance metric learning for image ranking. *IEEE Transactions on Neural Networks and Learning Systems 26*, 7 (2015), 1551–1559. 10

[LLMC17] LIU W., LIU X., MA H., CHENG P.: Beyond human-level license plate super-resolution with progressive vehicle search and domain priori GAN. In *Proceedings of the 2017 ACM on Multimedia Conference, MM 2017, Mountain View, CA, USA, October 23-27, 2017* (2017), pp. 1618–1626. 1

[LTH*17] LEDIG C., THEIS L., HUSZAR F., CABALLERO J., CUNNINGHAM A., ACOSTA A., AITKEN A. P., TEJANI A., TOTZ J., WANG Z., SHI W.: Photo-realistic single image super-resolution using a generative adversarial network. In *2017 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, Honolulu, HI, USA, July 21-26, 2017* (2017), pp. 105–114. 2, 3, 10

[LWD*17] LI C., WANG X., DONG W., YAN J., LIU Q., ZHA H.: Joint active learning with feature selection via cur matrix decomposition. *IEEE Transactions on Pattern Analysis and Machine Intelligence PP*, 99 (2017), 1–1. 10

[MFTM01] MARTIN D. R., FOWLKES C. C., TAL D., MALIK J.: A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In *ICCV* (2001), pp. 416–425. 8

[MLX*17] MAO X., LI Q., XIE H., LAU R. Y. K., WANG Z., SMOLLEY S. P.: Least squares generative adversarial networks. In *IEEE International Conference on Computer Vision, ICCV 2017, Venice, Italy, October 22-29, 2017* (2017), pp. 2813–2821. 3

[MMCS11] MASCI J., MEIER U., CIRESAN D. C., SCHMIDHUBER J.: Stacked convolutional auto-encoders for hierarchical feature extraction. In *Artificial Neural Networks and Machine Learning - ICANN 2011 - 21st International Conference on Artificial Neural Networks, Espoo, Finland, June 14-17, 2011, Proceedings, Part I* (2011), pp. 52–59. 3

[NCT16] NOWOZIN S., CSEKE B., TOMIOKA R.: f-gan: Training generative neural samplers using variational divergence minimization. In *Advances in Neural Information Processing Systems 29: Annual Conference on Neural Information Processing Systems 2016, December 5-10, 2016, Barcelona, Spain* (2016), pp. 271–279. 3

[RMC15] RADFORD A., METZ L., CHINTALA S.: Unsupervised representation learning with deep convolutional generative adversarial networks. *arXiv preprint arXiv:1511.06434* (2015). 3

[SCH*16] SHI W., CABALLERO J., HUSZAR F., TOTZ J., AITKEN A. P., BISHOP R., RUECKERT D., WANG Z.: Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network. In *2016 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2016, Las Vegas, NV, USA, June 27-30, 2016* (2016), pp. 1874–1883. 2

[SLJ*15] SZEGEDY C., LIU W., JIA Y., SERMANET P., REED S. E., ANGUELOV D., ERHAN D., VANHOUCKE V., RABINOVICH A.: Going deeper with convolutions. In *IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2015, Boston, MA, USA, June 7-12, 2015* (2015), pp. 1–9. 2

[SSH17] SAJJADI M. S. M., SCHÖLKOPF B., HIRSCH M.: Enhancenet: Single image super-resolution through automated texture synthesis. In *IEEE International Conference on Computer Vision, ICCV 2017, Venice, Italy, October 22-29, 2017* (2017), pp. 4501–4510. 3

[SZ14] SIMONYAN K., ZISSERMAN A.: Very deep convolutional networks for large-scale image recognition. *CoRR abs/1409.1556* (2014). arXiv:1409.1556. 2

[TLD*14] TRINH D. H., LUONG M., DIBOS F., ROCCHISANI J., PHAM C. D., NGUYEN T. Q.: Novel example-based method for super-resolution and denoising of medical images. *IEEE Trans. Image Processing 23*, 4 (2014), 1882–1895. 1

[TLLG17] TONG T., LI G., LIU X., GAO Q.: Image super-resolution using dense skip connections. In *IEEE International Conference on Computer Vision, ICCV 2017, Venice, Italy, October 22-29, 2017* (2017), pp. 4809–4817. 2, 4, 6, 8

[TSG13] TIMOFTE R., SMET V. D., GOOL L. J. V.: Anchored neighborhood regression for fast example-based super-resolution. In *IEEE International Conference on Computer Vision, ICCV 2013, Sydney, Australia, December 1-8, 2013* (2013), pp. 1920–1927. 2

[TSG14] TIMOFTE R., SMET V. D., GOOL L. J. V.: A+: adjusted anchored neighborhood regression for fast super-resolution. In *Computer Vision - ACCV 2014 - 12th Asian Conference on Computer Vision, Singapore, Singapore, November 1-5, 2014, Revised Selected Papers, Part IV* (2014), pp. 111–126. 6, 8

[TYL17] TAI Y., YANG J., LIU X.: Image super-resolution via deep recursive residual network. In *2017 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, Honolulu, HI, USA, July 21-26, 2017* (2017), pp. 2790–2798. 2, 6, 8

[TYLX17] TAI Y., YANG J., LIU X., XU C.: Memnet: A persistent memory network for image restoration. In *IEEE International Conference on Computer Vision, ICCV 2017, Venice, Italy, October 22-29, 2017* (2017), pp. 4549–4557. 6, 8

[WLX17] WANG J., LIU B., XU K.: Semantic segmentation of high-resolution images. *SCIENCE CHINA Information Sciences 60*, 12 (2017), 123101:1–123101:6. 10

[WLY*15] WANG Z., LIU D., YANG J., HAN W., HUANG T. S.: Deep networks for image super-resolution with sparse prior. In *2015 IEEE International Conference on Computer Vision, ICCV 2015, Santiago, Chile, December 7-13, 2015* (2015), pp. 370–378. 6, 8

[WXH17] WU X., XU K., HALL P.: A survey of image synthesis and editing with generative adversarial networks. *Tsinghua Science and Technology 22*, 6 (2017), 660–674. 3

[WYDCL18] WANG X., YU K., DONG C., CHANGE LOY C.: Recovering realistic texture in image super-resolution by deep spatial feature transform. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (June 2018). 1

[YLC13] YANG J., LIN Z., COHEN S.: Fast image super-resolution based on in-place example regression. In *2013 IEEE Conference on Computer Vision and Pattern Recognition, Portland, OR, USA, June 23-28, 2013* (2013), pp. 1059–1066. 1, 2

[YWHM08] YANG J., WRIGHT J., HUANG T. S., MA Y.: Image super-resolution as sparse representation of raw image patches. In *2008 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2008), 24-26 June 2008, Anchorage, Alaska, USA* (2008). 2

[YWHM10] YANG J., WRIGHT J., HUANG T. S., MA Y.: Image super-resolution via sparse representation. *IEEE Trans. Image Processing 19*, 11 (2010), 2861–2873. 2

[ZEP10] ZEYDE R., ELAD M., PROTTER M.: On single image scale-up using sparse-representations. In *Curves and Surfaces - 7th International Conference, Avignon, France, June 24-30, 2010, Revised Selected Papers* (2010), pp. 711–730. 8

[Zha16] ZHANG X.: Simultaneous optimization for robust correlation estimation in partially observed social network. *Neurocomputing 205* (2016), 455–462. 10

[ZLW*14] ZHU X., LIU J., WANG J., LI C., LU H.: Sparse representation for robust abnormality detection in crowded scenes. *Pattern Recognition 47*, 5 (2014), 1791–1799. 2

[ZML16] ZHAO J., MATHIEU M., LECUN Y.: Energy-based generative adversarial network. *arXiv preprint arXiv:1609.03126* (2016). 3

[ZPIE17] ZHU J., PARK T., ISOLA P., EFROS A. A.: Unpaired image-to-image translation using cycle-consistent adversarial networks. In *IEEE International Conference on Computer Vision, ICCV 2017, Venice, Italy, October 22-29, 2017* (2017), pp. 2242–2251. 3

[ZXZ*16] ZHANG C., XUE Z., ZHU X., WANG H., HUANG Q., TIAN Q.: Boosted random contextual semantic space based representation for visual recognition. *Inf. Sci. 369* (2016), 160–170. 10