

Improved Residual Dense Network for Large Scale Super-Resolution via Generative Adversarial Network

Inad A. Aljarrah and Eman M. Alshare

Department of Computer Engineering, Jordan University of Science and Technology, Irbid 22110, Jordan

Abstract: Recent single image super resolution (SISR) studies were conducted extensively on small upscaling factors such as x2 and x4 on remote sensing images, while less work was conducted on large factors such as the factor x8 and x16. Owing to the high performance of the generative adversarial networks (GANs), in this paper, two GAN's frameworks are implemented to study the SISR on the residual remote sensing image with large magnification under x8 scale factor, which is still lacking acceptable results. This work proposes a modified version of the residual dense network (RDN) and then it been implemented within GAN framework which named RDGAN. The second GAN framework has been built based on the densely sampled super resolution network (DSSR) and we named DSGAN. The used loss function for the training employs the adversarial, mean squared error (MSE) and the perceptual loss derived from the VGG19 model. We optimize the training by using Adam for number of epochs then switching to the SGD optimizer. We validate the frameworks on the proposed dataset of this work and other three remote sensing datasets: the UC Merced, WHU-RS19 and RSSCN7. To validate the frameworks, we use the following image quality assessment metrics: the PSNR and the SSIM on the RGB and the Y channel and the MSE. The RDGAN evaluation values on the proposed dataset were 26.02, 0.704, and 257.70 for PSNR, SSIM and the MSE, respectively, and the DSGAN evaluation on the same dataset yielded 26.13, 0.708 and 251.89 for the PSNR, the SSIM, and the MSE.

Keywords: single image super-resolution, remote sensing, generative adversarial network, residual dense network, residual dense generative adversarial network.

1. Introduction

Many computer vision systems involve having a high-quality image that eases extracting important features and employing this information to achieve some purposes. In image processing the task of reconstructing high-resolution (HR) image from low resolution (LR) one is called image super resolution (SR). The importance of the SR task has been shown in various tasks such as automated license plate [1] and face recognition [2], it also has been used in medical diagnostic [3] framework with the intent to maximize the performance of diagnosing illnesses and ensuring pathological invariant.

The image SR system aims to reconstruct an HR image of LR input, by extracting most salient features, mapping these LR features to the HR space and generating an HR image. SR task is considered as inverse and ill-posed problem that has no unique solution which can treat all image domains and tackles all challenges that arise in the LR input image (e.g., modeling diverse levels of noise, blur kernel and upscale factor). Abundance of methods has been investigated to handle the SR problem in many computer vision systems appropriately. The most widely used approach is the Interpolation-based methods [4] with diverse settings (e.g.,

Bilinear, Bicubic, and Nearest Neighbor for pixel imputation). These methods have been widely used due to their real-time performance, however, the drawback of these methods, is that they treat all kinds of domains in the same way and ignores the fact of noise present and the nature of the input image, hence it failed in retrieving high-frequency details with high visual quality.

Image super resolution studies are divided into two taxonomies regarding the inputs, single image super resolution (SISR) [5] and multi-image super-resolution (MISR) [6]. The MISR type concerned with having multiple LR images to reconstruct an HR image, i.e., employs or relies on image registration to allow extracting details that cannot be retrieved using a single image, hence, return a high-quality image. SISR methods take a single LR image to recover an HR image using the most presented details in the input. While multi-image input can yield nearly free noise images using more than one image, however, lacking a multi-images dataset and computation complexity restrict those methods. As a result, the researchers sought for another solution based on the available data and switched to the SISR methods that use a single LR image to reconstruct an HR image.

Recent research of SISR using convolution neural networks (CNNs) and generative adversarial networks (GANs) have shown good performance. Pioneer work of CNN models is the SRCNN [5]. SRCNN model consists of three layers: feature extraction layer, non-linear mapping, and reconstruction layer. This network learns an end-to-end mapping between LR and HR images which outperformed the state-of-the-art methods. Further improvements had been conducted on the SRCNN to speed up the training in [7]. Later, many works had been designed based on the generative models such as the GANs model. The first work on SR using GAN is the SRGAN [8] model. SRGAN model employed the residual skip connection proposed in [9] as the generator network and design a classifier network as the discriminator network. Unlike previous work the SRGAN model used perceptual loss that derived from the difference between the VGG feature maps of the LR and HR images.

Many works are conducted on the remote sensing image [10] to recover an HR image from the available LR image using learning-based methods. Since acquiring an HR satellite image is difficult due to many factors such as the weather condition and position of the capturing sensor and image sensor performance, in addition to the motion blurs, all these factors affect the quality of the image and yield a noisy degraded image. As a result, HR satellite image demand in computer vision systems has led researchers, specifically, to study this topic immensely.

In this work, we worked on developing a GAN model based on a single input image, to restore a realistic HR image with large factor (x8) from the observed LR image. We will explore a new solution by employing GAN model with appropriate generator-discriminator topology to acquire high vitality images.

The contributions of this work are proposing and conducting the work on a new remote sensing dataset of residential area that suffers from noise-contaminated imaging conditions. Also, developing two GAN frameworks, RDGAN and DSGAN and improving the RDN model on large scale factor (x8). The remainder of the paper is arranged as follows. Section 2 gives an overview of the SR-related work. Next, section 3 shows the proposed methodology in this work in details, and then it comes up with the training settings and the used datasets. After that, section 4 shows the results in detail. Finally, section 5 shows ablation study on the RDGAN model. Finally, section 6 concludes the work and presents a future work.

2. Related work

The emerging SISR deep learning models had been trained in supervised manners using paired LR/HR images. In this section, we take an overview of the CNN and GAN-based SR models. Early works in SISR adopted pre-up-sampling methods [11-13], which interpolated the LR image to the desired size and fed it to the model. However, these methods had trained in the high-resolution feature space which had increased the needed computation and memory space for training. Post-up-sampling methods solve these issues by replacing pre-up-sampling with up-sampling layers at the end of the models. An improvement on the SRCNN model has been made in [7]. This work has reformulated the SRCNN and introduced a deconvolutional layer at the end of the model. Furthermore, progressive up-sampling [14] and iterative up-and-down sampling [15] paradigms have been introduced into SISR models.

Many works had adopted skip connections to solve the vanishing/exploding gradient problem presented in the very deep CNN models. For instance, the work in [16] proposed an encoder-decoder framework with symmetric skip connections to tackle image restoration tasks, including denoising and SISR tasks with variance levels of noises. Also, the proposed VDSR in [17] had introduced global residual learning between the input image and output HR image to refine the high-frequency texture.

Furthermore, the deep residual channel attention network (RCAN) [18] explored residual channel attention and presented a residual in residual (RIR) design that used long and short skip connections to allow increasing network depth and resolves deep network complexity. Moreover, the work in [19] incorporated attention mechanisms with the residual learning for SISR of remote sensing images, and in [20] a single and multi-scale SR models were proposed: EDSR and MDSR, which enhanced the residual block by removing batch normalization layers. Another work in [21], which designed memory network (MemNet) that consists of memory blocks that determine which information to store for the restoration process and prevent information loss as network depth increases.

In [22], the authors had introduced a dual-path block that employed residual and dense connections. This design allows feature reusability and self and spatial attention blocks to handle the correlation among features at different levels. On

the other hand, Jiang et al. [23] had invented a hierarchical dense recursive network (HDRN) to improve the features reusing alongside memory usage. This facilitates integrating multi-scale features generated by different layers and decreases the power to compute gradients.

Work in [24] considered the SR problem as an inpainting task and proposed two-stage inpainting model concerns with structural and textural inpainting. This model reinforces constructing high-quality images by relying on edge information, however, it needs more computation to reconstruct the final HR. Super-resolution discriminative dense network SRDDN [25] is another introduced approach that aims to solve the drawback of treating all extracted features equally by introducing aggregation modules.

Wang et al. [26] had developed an EEDS network that uses shallow CNN to maintain global image structure and deep CNN for retrieving the high-frequency details, which speed up the training convergence. Another design in [27] has implemented convolutional layers with 1D and 2D convolution kernels, e.g., 5x1 and 3x3, to perform orientation-aware features extraction. In [28], the proposed model has been built based on multi-scale feature extractions using a receptive field block (RFB), and to minimize the feature map size, 1x1 convolution layers were employed.

In the earliest studies of SISR using GAN, SRGAN [8] adopted residual block with content loss based on the extracted high-level feature from the pre-trained VGG-19 network [29, 30], it also used Parametric ReLU [31] activation unit at the generator rather than ReLU activation. Further refinement on residual blocks introduced in [32], and residual-in-residual dense block (RRDB) has been proposed to enhance GAN-generated images. Contrary to the SRGAN, leaky rectified activation unit (LeakyReLU) [33] and Relativistic Discriminator [34] were used in this work. Another improvement on the SRGAN discriminator model has been conducted in [35] to compensate for loss information and, preserving local details for accurate classification.

Work in [10] proposed a dense residual generative adversarial network (DRGAN). This work used dense-residual units (DRUs) and residual learning at the generator and adopted discriminator training according to the WGAN-GP [36] for stability. Further improvement in [37] is implemented, which had designed EEGAN with an edge-enhancement method to remove artifacts of the previous GAN models. Also, the proposed model in [38] has adopted a structure-preserving SR method (SPSR) to recover image structures and sharpen edges by adopting gradient guidance and introducing a new gradient loss.

3. Methodologies

To solve the SR task on remote sensing images with large scale factor, we implemented two GAN frameworks that consist of a generator and discriminator networks. In this section we will illustrate the proposed residual dense generative adversarial network (RDGAN) and the dense sampling generative adversarial network (DSGAN).

3.1. Residual Dense GAN

The residual dense GAN (RDGAN) model consists of two networks: generator, and discriminator networks. The generator is configured as a modified version of the residual dense network (RDN) proposed in [39] to reconstruct high resolution images (HR). For the discriminator network, we

proposed a new model with small capacity, unlike most of the recent proposed GAN models.

Generator network as shown in Figure.1, the modified RDN model consists of external and shallow feature extraction layers, residual dense block (RDB), dense feature fusion and ended by up-sampling layers. Let's denote I^{LR} and I^{SR} as the input and the output of the model. The external and the shallow convolution layers extract features from the input I^{LR} .

$$F_{\text{external}} = \text{Conv2D}_{\text{external}}(I^{LR})$$

$$F_{\text{shallow}} = \text{Conv2D}_{\text{shallow}}(I^{LR})$$

Where $\text{Conv2D}_{\text{external}}$ and $\text{Conv2D}_{\text{shallow}}$ denote the external and the shallow convolution layers. The output F_{external} is used for global residual learning and F_{shallow} is used as input to RDB. Suppose we have N residual dense blocks (RDNs), the output F_n of the n-th RDB can be obtained by

$$F_n = \text{RDB}_n(F_{n-1})$$

$$F_n = \text{RDB}_n(\text{RDB}_{n-1}(\dots(\text{RDB}_1(F_{\text{shallow}}))\dots))$$

Where RDB_n denotes the operations of the nth-RDB. The RDB is a composite of convolution layers, rectified linear units (ReLU), local residual learning and dense feature fusion. 3.3, shows the used RDB design in this work. Assume that F_{n-1} is the input to the n-th RDB then we can formulate the output of the n-th RDB as:

$$F_{\text{conv-1}} = \text{Conv2D}(F_{n-1})$$

$$F_{\text{conv-2}} = \text{Conv2D}([F_{\text{conv-1}}, F_{n-1}])$$

$$F_{\text{conv-c}} = \text{Conv2D}([F_{n-1}, F_{\text{conv-1}}, F_{\text{conv-2}}, \dots, F_{\text{conv-c-1}}])$$

$$F_{\text{LDFF}} = \text{Concat}([F_{n-1}, F_{\text{conv-1}}, F_{\text{conv-2}}, \dots, F_{\text{conv-c}}])$$

$$F_{\text{conv1x1}} = \text{Conv2D}_{1x1}(F_{\text{LDFF}})$$

$$F_{\text{LRL}} = F_{\text{conv1x1}} + F_{n-1}$$

Where c represents the number of convolution layers in the RDB, the Conv2D represent the convolutional operation and Concat represents the dense feature fusion.

Sequence of RDBs is used to extract hierarchical features then these features are passed to dense feature fusion (DFF) operation and followed by global residual learning (GRL) connection. The feature maps from the external convolution layer and the output of the dense feature fusion are fed to sum operation that represents the global residual learning.

$$F_{\text{GDFD}} = \text{Conv2D}_{1x1}(\text{Concat}([F_1, F_2, \dots, F_{n-1}, F_n]))$$

$$F_{\text{GRL}} = F_{\text{GDFD}} + F_{\text{external}}$$

The residual dense block as shown in Figure.2 consists of a sequence of convolutional layers with local dense feature (LDFF) fusion and local residual learning (LRL). At the end of the generator, the sequence of upscaling modules is used to up-sample the output image to the target size. The upscaling module consists of convolution layer, shuffle operation and ReLU activation layer as shown in Figure.3.

The discriminator network consists of sequential convolutional layers followed by batch normalization and Leaky ReLU except for the first convolution layer where the batch normalization is omitted. Then flatten, dense and batch normalization layers used to output one value that implies if the input image is an HR or SR image. The discriminator

scheme is shown in Fig.4. All convolutional layers used 3x3 kernel size while filters started from 64 and doubled until the last layer and every layer used stride of 2. Unlike SRGAN discriminator network, we omitted the fully connected layer in the proposed discriminator.

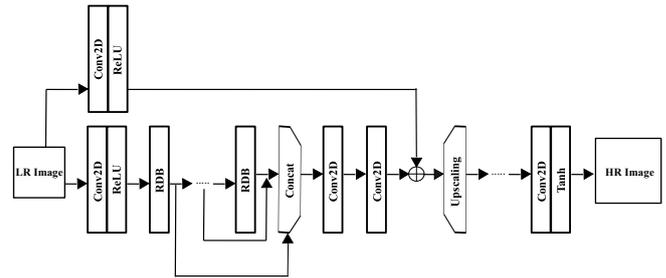


Figure 1. RDGAN's generator network design.

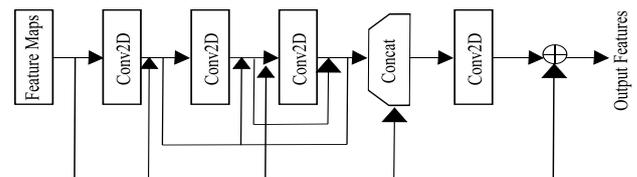


Figure 2. Residual dense block (RDB) design.

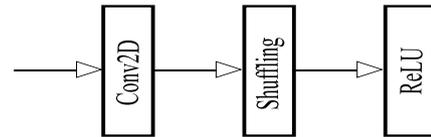


Figure 3. Upscaling block design.

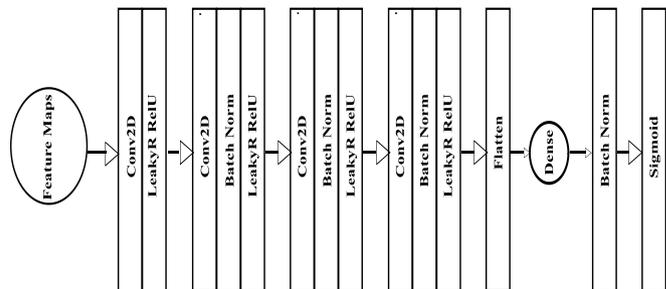


Figure 4. RDGAN's discriminator network.

3.2 Dense Sampling GAN

In this framework, we employ the proposed dense-sampling super-resolution network (DSSR) method in [40] as a generator network. Similar to the RDGAN framework, this model adopts local residual connections and global dense connections within the generator network. The discriminator network is similar to the RDGAN with little changes. In the next subsections, we will describe the proposed generator and discriminator of the DSGAN framework.

We used the same discrimination of the RDGAN model; however, we used two fully connected layers and removed the batch normalization layer at the end of the model before the final activation layer. Also, we use large capacity discriminator with more layers.

3.3. Dataset

The experiments are conducted on a collected dataset from Google earth API that contains 7000 images with resolution 4800x2400. The used datasets for training contain 1400 and 800 images. Due to resource limitation on GPUs and storage,

we used 1400 images for the training the RDGAN model and 700 for training the DSGAN model, next we cropped sub-images from every image then randomly flip the patches horizontally or vertically, also we set the batch size to 8 yielding training datasets of 11200 and 6400 images. The reset 200 and 100 images are used to validate the models, in addition we used three other datasets: UC Merced [41], WHU-RS19 [42] and RSSCN7 [43].

3.4. Training Details

For preparing the input of the models, we imported the images in BGR format and normalized the images to range [0,1], then randomly cropped 8 patches of size 256x256 from every image to form the HR images set, next we down sample these images using bicubic interpolation method by the required upscale factor, which is set to x8 in this work. The down-sampled image is fed to the model as an input. The training batch size also is set to 8 in all conducted experiments.

For both models: RDGAN (Residual dense GAN) and DSGAN (Dense sampling GAN), we adopt Adam [46] for optimizing the models, the learning rate set to 0.0001 while we left β_1 and β_2 with the default values. The weight of MSE, perceptual and the adversarial loss are set to 1, 0.00001 and 0.00001, respectively. The perceptual loss is driven from the output features that are taken from the 20th layer of the pretrained visual geometry group (VGG19) network before the activation layer.

For all convolutional layers in the two generators of the proposed models, except in the upscaling module of the RDGAN model, the growth rate is set to 64 and for the upscaling module the growth rate is set to 256. For all convolutional layers in both generators and discriminators the used kernel size is 3x3 and same padding. Generator architecture in the RDGAN model consists of 18 residual dense blocks (RDB) and 3 convolution layers within the RDB.

4. Results

To validate the work, we used PSNR, SSIM and MSE metrics to assess model performance. We tested the models on our dataset, and three other remote sensing datasets compared with previous work. In this study we selected the results of bicubic interpolation and the DSSR model as a baseline. Also, we presented sample of generated images from the two proposed models to assess the quality of the images.

4.1. Reference Quantitative Evaluation

First, we tested both models on our collected dataset and compared the result with bicubic interpolation implemented in OpenCV. Table 1 shows the evaluated objective metrics on our proposed dataset of 2400 images using the proposed models.

The peak signal-to-noise ratio (PSNR) and the structural similarity (SSIM) are evaluated on both RGB and the Y luminance channel of transformed YCbCr space while the mean squared error (MSE) is evaluated on RGB channels only, higher PSNR and SSIM values indicate better quality. We evaluated PSNR and SSIM on Y channel similar to [40,44] for comparison. The PSNR and SSIM indicate the average value on the RGB channels while the PSNR_Y and SSIM_Y indicate the average on the Y channel. Highest

value for PSNR and SSIM indicates better quality while lowest value of MSE metric indicates better quality.

Table 1 shows the quantity values on the proposed dataset of the DSGAN and RDGAN frameworks and compares between the proposed models and bicubic interpolation and the state-of-the-art DSSR [40] after we trained the model on the same data set. As shown both models perform well on the test samples, and the RDGAN boost the RDN model by 0.21, 0.009, 0.21, 0.008 and 13.35 in PSNR, SSIM, PSNR_Y, SSIM_Y, and MSE, respectively. The DSGAN boost the DSSR model by 0.03, 0.001, 0.03, 0.001 and 2.14 in PSNR, SSIM, PSNR_Y, SSIM_Y, and MSE, respectively.

Table 1. Average PSNR, SSIM on RGB and Y channel and MSE of Bicubic method, DSSR, RDN, RDGAN and DSGAN models on the proposed test set on 2400 images. The bold and underlined value indicates the best result.

Model / Metric	PSNR	SSIM	PSNR_Y	SSIM_Y	MSE
Bicubic	22.37	0.559	23.69	0.601	434.50
DSSR	24.76	0.673	26.10	0.707	254.03
RDN	24.47	0.660	25.81	0.696	271.05
RDGAN	24.68	0.669	26.02	0.704	257.70
DSGAN	24.79	0.674	26.13	0.708	251.89

To compare with other works, we used MHAN [44] and tested our models on the used dataset in those works. as shown in Table 2 the evaluation results of MHAN and DSSR are taken based on different sample than the samples used to validate the proposed models. We compared the bicubic interpolation with the proposed models. Comparing with the proposed models that is trained on residential remote sensing images while the work in [44] are trained on various types of remote sensing, we find that our models are promised and can perform well on unseen data.

It can be observed from Table 2 that the RDGAN model outperforms the MHAN model in SSIM with difference 0.007 on the RSSCN7 dataset. The RDGAN also outperforms the MHAN in the SSIM of the WHU-RS19 dataset with increment by 0.006. Another comparison with the MHAN is conducted on the proposed Test30 dataset in [44]. In Table 3 the average of the PSNR, SSIM and MSE, shows that the proposed RDGAN model has a better result in the SSIM metric by an amount of 0.002.

Table 2. Average PSNR and SSIM on the Y channel of Bicubic method, MHAN, RDGAN and DSGAN models on the RSSCN7 and WHU-RS19 dataset. The bold and underlined value indicates the best result.

	<u>Bicubic</u> PSNR/SS IM	<u>MHAN</u> [44] PSNR/SS IM	<u>DSGAN</u> PSNR/SS IM	<u>RDGAN</u> PSNR/SS IM
RSSCN 7	25.25 / 0.544	26.34 /0.588	26.34 / 0.596	26.35 / 0.596
WHU- RS19	25.30 / 0.589	27.03 / 0.646	26.87/ 0.653	26.79 / 0.652

Table 3. Average PSNR, SSIM and MSE of Bicubic method, RDGAN and DSGAN models on the proposed Test30 dataset in [44]. The bold and underlined value indicates the best result.

	PSNR	SSIM	PSNR_Y	SSIM_Y
Bicubic	25.10	0.535	26.45	0.576
RDGAN	26.28	0.588	27.66	0.625
DSGAN	25.89	0.575	27.63	0.624
MHAN	—	—	27.83	0.623

In Table 4, Table 5 and Table 6, the PSNR, SSIM were evaluated on each class in the WHU-RS19, UC Merced and RSSCN7 and compared the results with bicubic method and MHAN. The results show that the proposed models are promising even though the proposed models are trained only on a residential area.

Table 4. Average PSNR and SSIM on Y channel of each class in the WHU-RS19 dataset for each method.

WHU-RS19	Bicubic	MHAN [44]	RDGAN	DSGAN
	PSNR/SSIM	PSNR/SSIM	PSNR/SSIM	PSNR/SSIM
Airport	22.76/0.555	24.64/0.633	24.48/0.641	24.52/0.643
Beach	37.68/0.940	41.42/0.948	37.79/0.947	38.91/0.947
Bridge	27.74/0.805	30.51/0.840	30.60/0.843	30.67/0.844
Commercial	20.31/0.417	21.56/0.501	21.71/0.512	21.77/0.514
Desert	36.45/0.896	38.17/0.903	37.84/0.901	37.51/0.901
Farmland	31.72/0.769	33.20/0.793	33.15/0.796	33.11/0.796
Football Field	22.36/0.590	24.99/0.668	24.98/0.677	25.12/0.680
Forest	24.79/0.407	25.58/0.452	25.45/0.450	25.52/0.450
Industrial	21.26/0.473	23.20/0.557	23.47/0.584	23.49/0.585
Meadow	32.97/0.786	34.69/0.806	34.24/0.798	34.16/0.798
Mountain	21.98/0.335	22.73/0.394	22.64/0.394	22.64/0.393
Park	23.92/0.513	25.18/0.568	25.01/0.569	25.02/0.570
Parking	20.97/0.556	22.43/0.629	22.54/0.648	22.62/0.650
Pond	27.46/0.744	29.15/0.785	28.89/0.773	28.93/0.779
Port	21.97/0.631	23.63/0.702	23.77/0.704	23.85/0.709
Railway Station	21.89/0.409	23.04/0.479	22.98/0.488	23.00/0.489
Residential	19.42/0.420	21.0/0.513	21.17/0.538	21.18/0.537
River	24.59/0.539	25.88/0.585	25.89/0.592	25.97/0.595
Viaduct	20.54/0.409	22.55/0.514	22.41/0.529	22.47/0.529
Average	25.30/0.589	27.03/0.646	26.79/0.652	26.87/0.653

Table 5. Average PSNR, SSIM on Y channel of each class in the UC Merced test dataset for each method.

UC Merced	Bicubic	RDGAN	DSGAN
	PSNR/SSIM	PSNR/SSIM	PSNR/SSIM
Agricultural	19.29/0.325	21.39/0.348	22.59/0.398
Airplane	19.06/0.657	23.31/0.722	25.16/0.779
Baseball diamond	22.28/0.739	26.34/0.766	29.05/0.814
Beach	21.22/0.742	25.30/0.758	27.58/0.815
Buildings	18.23/0.582	22.31/0.676	24.28/0.747
Chaparral	17.51/0.423	20.46/0.491	21.23/0.551
Dense residential	17.82/0.545	21.52/0.631	23.50/0.714
Forest	18.71/0.397	20.96/0.426	22.67/0.487
Freeway	18.72/0.599	23.34/0.675	26.02/0.750
Golf course	21.93/0.738	26.62/0.761	29.38/0.811
Harbor	16.49/0.540	19.01/0.615	20.58/0.684
Intersection	19.08/0.584	22.69/0.644	24.67/0.705
Medium residential	18.72/0.532	22.20/0.609	23.86/0.667
Mobile home parks	17.29/0.474	20.47/0.582	21.55/0.635
Overpass	18.47/0.547	22.15/0.620	24.55/0.689
Parking lot	16.50/0.471	18.93/0.549	20.05/0.617
River	20.42/0.637	24.20/0.663	26.99/0.718
Runway	19.63/0.651	24.41/0.710	27.57/0.780
Sparse residential	19.28/0.558	23.02/0.602	25.32/0.668
Storage tanks	18.65/0.628	23.08/0.700	25.14/0.759
Tennis court	19.49/0.631	23.75/0.675	26.46/0.743
Average	18.99/0.571	22.64/0.630	24.68/0.692

Table 6. Average PSNR and SSIM on Y channel of each class in the RSSCN7 dataset for each method.

RSSCN7	Bicubic	MHAN [44]	RDGAN	DSGAN
	PSNR/SSIM	PSNR/SSIM	PSNR/SSIM	PSNR/SSIM
Grass	30.32/0.729	31.64/0.752	31.46/0.751	31.42/0.751
Industry	21.31/0.454	22.66/0.524	22.85/0.541	22.85/0.541
River lake	27.09/0.695	28.42/0.731	28.31/0.732	28.30/0.732
Filed	30.37/0.676	31.38/0.691	31.37/0.696	31.34/0.696
Forest	25.58/0.423	26.17/0.458	26.15/0.461	26.14/0.460
Resident	20.55/0.386	21.63/0.460	21.69/0.475	21.69/0.474
Parking	21.53/0.446	22.48/0.497	22.65/0.515	22.65/0.515
Average	25.25/0.544	26.34/0.588	26.35/0.596	26.34/0.596

4.2. Qualitative Evaluation

In this section, we provide random samples of the generated images of the proposed 2400 images were used in evaluation. In Fig 5, a comparison among the images represented the DSGAN, RDGAN, HR and bicubic images, the proposed models are outperforming the traditional method in generating images with high frequent textures.

4.3. Ablation Study

This section studies the effect of the local feature fusion part within the RDB and using the MSE loss with the adversarial loss without using the perceptual loss. In table 7, we assess all models on 2400 images to find the effect of the removed part. We found that the perceptual loss increases the assessment metrics with small amount while the concatenate layer in the RDB has a significant effect on the results. The concatenate layer has more effect in improving model's performance compared with the perceptual loss.

Table 7. Average PSNR, SSIM on the Y channel of RDGAN model with various settings on the proposed test set of 2400 images. The bold and underlined value indicates the best result.

	M-RDGAN	M-RDGAN	M-RDGAN
Perceptual Loss	X	√	√
Local Dense Feature Fusion	√	x	√
PSNR_Y / SSIM_Y	25.93/0.703	25.70/0.692	26.0195/0.704

5. Conclusion

This work proposed two GAN models to treat the SISR task under large scale factor (x8). First, we proposed an improved the version of the RDN model in [39] by adding an external feature extraction layer that improved the reconstructed image and then we employed the modified RDN in a GAN framework. Also, we highlighted the effect of the dense connection and the perceptual loss that is derived from the generated feature maps of the VGG19 network. The proposed model shows a good performance in both quality and quantity metrics on the proposed dataset. Also, the models show a comparable result to one of the recent works on three other datasets.

We also proposed the DSGAN framework based on the DSSR network [40] to boost the performance of the original model by building GAN based framework. We adopted the same setting of the DSSR as we went along with their findings regarding the best hyper parameters values. The comprehensive experimental results have demonstrated that this work returns comparable results with others with small network capacity.

For future work, we can study the performance of the proposed frameworks on various degradation models as explored in work [45], also derive perceptual loss function from the discriminator network rather than VGG19 model.

References

- [1] C. H. Chuang, L. W. Tsai, M. S. Deng, J. W. Hsieh and K. C. Fan, "Vehicle license plate recognition using super-resolution technique," 2014 11th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS), pp.411-416, 2014: IEEE.
- [2] G. Gao, D. Zhu, M. Yang, H. Lu, W. Yang and H. Gao, "Face image super-resolution with pose via nuclear norm regularized structural orthogonal

procrustes regression,” *Neural Computing and Applications*, Vol. 32, No. 9, pp. 4361-4371, 2020.

[3] C. H. Pham, A. Ducournau, R. Fablet and F. Rousseau, “Brain MRI super-resolution using deep 3D convolutional networks,” *2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017)*, pp.197-200, 2017: IEEE.

[4] V. H. Patil and D. S. Bormane, “Interpolation for super resolution imaging,” *Innovations and Advanced Techniques in Computer and Information Sciences and Engineering: Springer*, pp.483- 489, 2007.

[5] W. Shi, J. Caballero, F. Huszár, J. Totz, AP. Aitken, R. Bishop, et al., “Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network,” *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp.1874-1883, 2016. https://cv-foundation.org/openaccess/content_cvpr_2016/papers/Shi_Real-Time_Single_Image_CVPR_2016_paper.pdf

[6] M. R. Arefin, V. Michalski, P. L. St-Charles, A. Kalaitzis, S. Kim, S. E. Kahou, et al., “Multi-image super-resolution for remote sensing using deep recurrent networks,” *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pp.206-207, 2020. https://openaccess.thecvf.com/content_CVPRW_2020/papers/w11/Arefin_Multi-Image_Super-Resolution_for_Remote_Sensing_Using_Deep_Recurrent_Networks_CVPRW_2020_paper.pdf

[7] C. Dong, C. C. Loy and X. Tang, “Accelerating the super-resolution convolutional neural network,” *European conference on computer vision*, pp.391-407, 2016: Springer.

[8] C. Ledig, L. Theis, F. Huszár, J. Caballero, A. Cunningham, A. Acosta, et al., “Photo-realistic single image super-resolution using a generative adversarial network,” *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp.4681-4690, 2017. https://openaccess.thecvf.com/content_cvpr_2017/papers/Ledig_Photo-Realistic_Single_Image_CVPR_2017_paper.pdf

[9] K. He, X. Zhang, S. Ren and J. Sun, “Deep residual learning for image recognition,” *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp.770-778, 2016. https://openaccess.thecvf.com/content_cvpr_2016/papers/He_Deep_Residual_Learning_CVPR_2016_paper.pdf

[10] W. Ma, Z. Pan, F. Yuan and B. Lei, “Super-resolution of remote sensing images via a dense residual generative adversarial network,” *Remote Sensing*, Vol. 11, No. 21, p.2578, 2019.

[11] J. Kim, J. K. Lee and K. M. Lee, “Deeply-recursive convolutional network for image super-resolution,” *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp.1637-1645, 2016. https://openaccess.thecvf.com/content_cvpr_2016/papers/Kim_Deeply-Recursive_Convolutional_Network_CVPR_2016_paper.pdf

[12] C. Dong, C. C. Loy, K. He and X. Tang, “Image super-resolution using deep convolutional networks,” *IEEE transactions on pattern analysis and machine intelligence*, Vol. 38, No. 2, pp.295-307, 2015. <https://arxiv.org/pdf/1501.00092.pdf>

[13] Y. Tai, J. Yang and X. Liu, “Image super-resolution via deep recursive residual network,” *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp.3147-3155, 2017. https://openaccess.thecvf.com/content_cvpr_2017/papers/Tai_Image_Super-Resolution_via_CVPR_2017_paper.pdf

[14] M. Zhao, X. Liu, H. Liu and K. K. L. Wong, “Super-resolution of cardiac magnetic resonance images using Laplacian pyramid based on generative adversarial networks,” *Computerized Medical Imaging and Graphics*, Vol. 80, p.101698, 2020.

[15] Y. Yu, X. Li and F. Liu, “E-DBPN: Enhanced deep back-projection networks for remote sensing scene image super resolution,” *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 58, No. 8, pp.5503- 5515, 2020.

[16] X. Mao, C. Shen and Y. B. Yang, “Image restoration using very deep convolutional encoder-decoder networks with symmetric skip connections,” *arXiv preprint arXiv:160309056*. 2016.

[17] X. Mao, C. Shen and Y. B. Yang, “Image restoration using very deep convolutional encoder-decoder networks with symmetric skip connections,” *Advances in neural information processing systems*, Vol. 29, pp.2802-2810, 2016. https://openaccess.thecvf.com/content_cvpr_2016/papers/Kim_Accurate_Image_Super-Resolution_CVPR_2016_paper.pdf

[18] Y. Zhang, K. Li, K. Li, L. Wang, B. Zhong and Y. Fu, “Image super-resolution using very deep residual channel attention networks,” *Proceedings of the European conference on computer vision (ECCV)*, pp.286-301, 2018. https://openaccess.thecvf.com/content_ECCV_2018/papers/Yulun_Zhang_Image_Super-Resolution_Using_ECCV_2018_paper.pdf

[19] J. M. Haut, R. Fernandez-Beltran, M. E. Paoletti, J. Plaza and A. Plaza, “Remote sensing image superresolution using deep residual channel attention,” *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 57, No. 11, pp.9277-9289, 2019.

[20] B. Lim, S. Son, H. Kim, S. Nah and K. Mu Lee, “Enhanced deep residual networks for single image super-resolution,” *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pp.136-144, 2017. https://openaccess.thecvf.com/content_cvpr_2017_workshops/w12/papers/Lim_Enhanced_Deep_Residual_CVPR_2017_paper.pdf

[21] Y. Tai, J. Yang, X. Liu and C. Xu, “Memnet: A persistent memory network for image restoration,” *Proceedings of the IEEE international conference on computer vision*, pp.4539-4547, 2017.

[22] D. W. Chen and C. H. Kuo, “Modified Dual Path Network With Transform Domain Data for

Image Super-Resolution,” IEEE Access, Vol. 8, pp.97975-97985, 2020.

[23] K. Jiang, Z. Wang, P. Yi and J. Jiang, “Hierarchical dense recursive network for image super-resolution,” Pattern Recognition, 107, p.107475, 2020.

[24] K. Nazeri, H. Thasarathan and M. Ebrahimi, “Edge-informed single image super-resolution,” Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops, 2019. https://openaccess.thecvf.com/content_ICCVW_2019/papers/AIM/Nazeri_Edge-Informed_Single_Image_Super-Resolution_ICCVW_2019_paper.pdf

[25] J. Ma, X. Wang and J. Jiang, “Image super resolution via dense discriminative network,” IEEE Transactions on Industrial Electronics, Vol. 67, No. 7, pp. 5687-5695, 2019.

[26] Y. Wang, L. Wang, H. Wang and P. Li, “End-to-end image super-resolution via deep and shallow convolutional networks,” IEEE Access, Vol. 7, pp.31959-31970, 2019.

[27] D. Chen, Z. He, Y. Cao, J. Yang, Y. Cao, M. Y. Yang, et al, “Deep Neural Network for Fast and Accurate Single Image Super-Resolution via Channel-Attention-based Fusion of Orientation-aware Features,” arXiv preprint arXiv:191204016, 2019. <https://arxiv.org/pdf/1912.04016.pdf>

[28] T. Shang, Q. Dai, S. Zhu, T. Yang and Y. Guo, “Perceptual extreme super-resolution network with receptive field block,” Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, pp.440-441, 2020. https://openaccess.thecvf.com/content_CVPRW_2020/papers/w31/Shang_Perceptual_Extreme_Super-Resolution_Network_With_Receptive_Field_Block_CVPRW_2020_paper.pdf

[29] J. Johnson, A. Alahi and L. Fei-Fei, “Perceptual losses for real-time style transfer and super-resolution,” European conference on computer vision, pp.694-711, 2016: Springer.

[30] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” arXiv preprint arXiv:14091556, 2014. [https://arxiv.org/pdf/1409.1556.pdf\(2014\).pdf](https://arxiv.org/pdf/1409.1556.pdf(2014).pdf)

[31] K. He, X. Zhang, S. Ren and J. Sun, “Delving deep into rectifiers: Surpassing human-level performance on imagenet classification,” Proceedings of the IEEE international conference on computer vision, pp.1026-1034, 2015. https://openaccess.thecvf.com/content_iccv_2015/papers/He_Delving_Deep_into_ICCV_2015_paper.pdf

[32] X. Wang, K. Yu, S. Wu, J. Gu, Y. Liu, C. Dong, et al., “Esrgan: Enhanced super-resolution generative adversarial networks,” Proceedings of the European Conference on Computer Vision (ECCV) Workshops, 2018. https://openaccess.thecvf.com/content_ECCVW_2018/papers/11133/Wang_ESRGAN_Enhanced_Super-Resolution_Generative_Adversarial_Networks_ECCVW_2018_paper.pdf

[33] B. Xu, N. Wang, T. Chen and M. Li, “Empirical evaluation of rectified activations in convolutional network,” arXiv preprint

arXiv:150500853, 2015.

<https://arxiv.org/pdf/1505.00853.pdf%E3%80%82ReLU>

[34] A. Jolicoeur-Martineau, “The relativistic discriminator: a key element missing from standard GAN,” arXiv preprint arXiv:180700734, 2018. <https://arxiv.org/pdf/1807.00734.pdf>

[35] D. Lee, S. Lee, H. Lee, K. Lee and H. J. Lee, “Resolution-preserving generative adversarial networks for image enhancement,” IEEE Access, Vol. 7, pp.110344-110357, 2019.

[36] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin and A. Courville, “Improved training of wasserstein gans,” arXiv preprint arXiv:170400028, 2017. <https://arxiv.org/pdf/1704.00028.pdf>

[37] K. Jiang, Z. Wang, P. Yi, G. Wang, T. Lu and J. Jiang, “Edge-enhanced GAN for remote sensing image superresolution,” IEEE Transactions on Geoscience and Remote Sensing, Vol. 57, No. 8, pp.5799-5812, 2019.

[38] C. Ma, Y. Rao, Y. Cheng, C. Chen, J. Lu and J. Zhou, “Structure-preserving super resolution with gradient guidance,” Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp.7769-7778, 2020. https://openaccess.thecvf.com/content_CVPR_2020/papers/Ma_Structure-Preserving_Super_Resolution_With_Gradient_Guidance_CVPR_2020_paper.pdf

[39] Y. Zhang, Y. Tian, Y. Kong, B. Zhong and Y. Fu, “Residual dense network for image super-resolution,” Proceedings of the IEEE conference on computer vision and pattern recognition, pp.2472-2481, 2018. https://openaccess.thecvf.com/content_cvpr_2018/papers/Zhang_Residual_Dense_Network_CVPR_2018_paper.pdf

[40] X. Dong, X. Sun, X. Jia, Z. Xi, L. Gao and B. Zhang, “Remote sensing image super-resolution using novel dense-sampling networks,” IEEE Transactions on Geoscience and Remote Sensing. Vol. 59, No. 2, pp.1618-1633, 2020.

[41] Y. Yang and S. Newsam, “Bag-of-visual-words and spatial extensions for land-use classification,” Proceedings of the 18th SIGSPATIAL international conference on advances in geographic information systems, pp.270-279, 2010. https://faculty.ucmerced.edu/snewsam/papers/Yang_ACMGIS10_BagOfVisualWords.pdf

[42] D. Dai and W. Yang, “Satellite image classification via two-layer sparse coding with biased image representation,” IEEE Geoscience and Remote Sensing Letters, Vol. 8, No. 1, pp.173-176, 2010. <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.703.6870&rep=rep1&type=pdf>

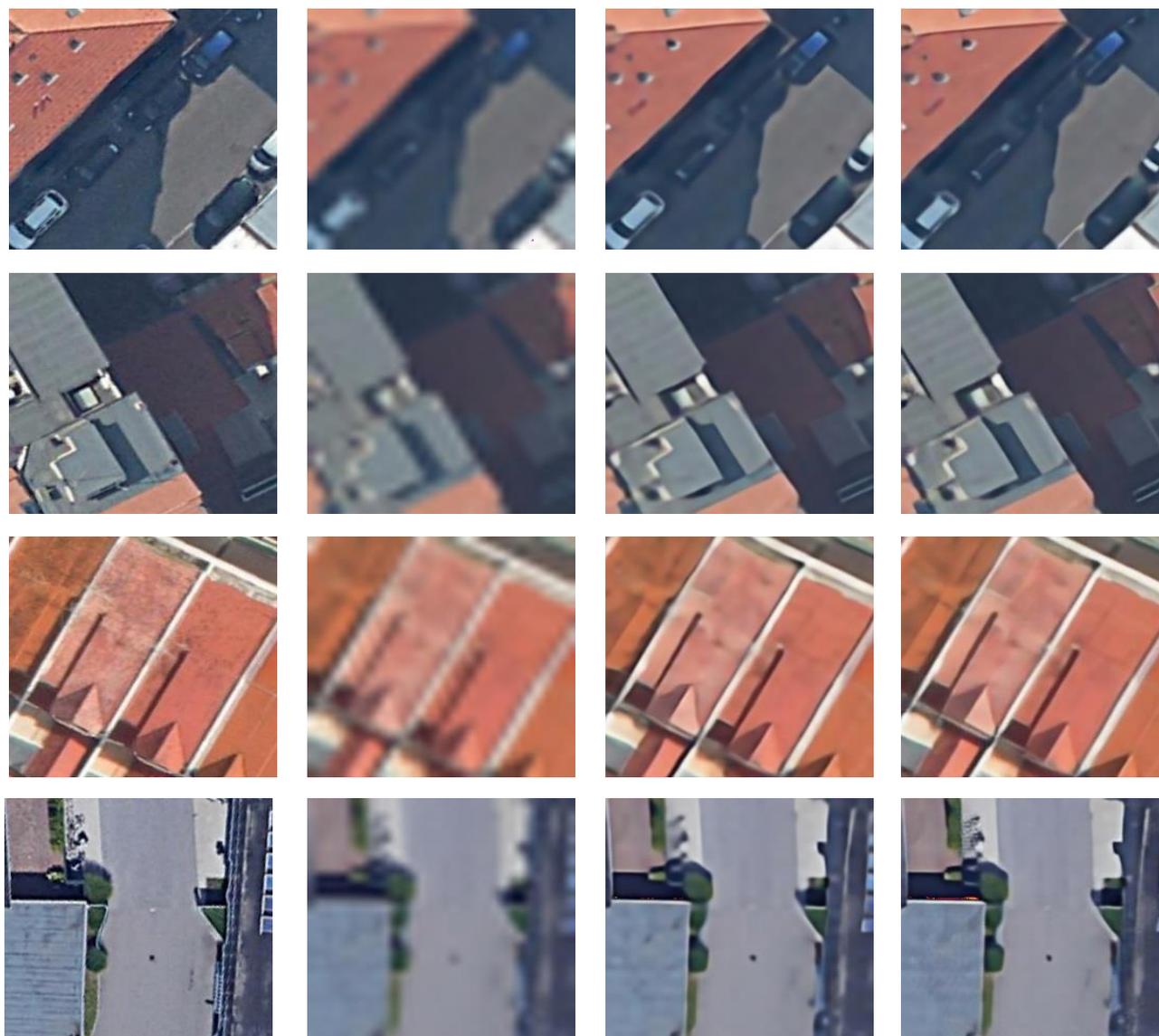
[43] Q. Zou, L. Ni, T. Zhang and Q. Wang, “Deep learning based feature selection for remote sensing scene classification,” IEEE Geoscience and Remote Sensing Letters, Vol. 12, No. 11, pp.2321-2325, 2015. http://mvr.whu.edu.cn/pubs/2015-IEEE_GRSL.pdf

[44] D. Zhang, J. Shao, X. Li and H. T. Shen, “Remote sensing image super-resolution via mixed high-order attention network,” IEEE Transactions on

Geoscience and Remote Sensing, Vol. 59, No. 6, pp.5183-5196, 2020.

[45] I. A. Aljarrah, "Effect of Image Degradation on Performance of Convolutional Neural Networks," International Journal of Communication Networks and Information Security, Vol. 13, No. 2, pp.215-219, 2021. <https://www.proquest.com/openview/a6b2ff5f9d04975992cebf6fac34ce35/1?pq-origsite=gscholar&cbl=52057>

[46] D. P. Kingma, J. Ba, "Adam: A method for stochastic optimization," arXiv preprint arXiv:14126980, 2014.



HR

Bicubic

RDGAN

DSGAN

Figure 5: Visual comparison of bicubic interpolation methods and our models on $\times 8$ factor