Coupled Adversarial Training for Remote Sensing Image Super-resolution

Sen Lei, Zhenwei Shi*, *Member IEEE*, and Zhengxia Zou*

***Abstract*—Generative adversarial network (GAN) has made great progress in recent natural image super-resolution tasks. The key to its success is the integration of a discriminator which is trained to classify whether the input is a real high-resolution (HR) image or a generated one. Arguably, learning a strong discriminative prior is essential for generating high-quality images. However, in remote sensing images, we discover through extensive statistical analysis that there are more low-frequency components than natural images, which may lead to a "discrimination-ambiguity" problem, i.e. the discriminator will become "confused" to tell whether its input is real or not when dealing with those low-frequency regions, and therefore, the quality of generated HR image may be deeply affected. To address this problem, we propose a novel GAN-based super-resolution algorithm named Coupled-Discriminated Generative Adversarial Networks (CDGAN) for remote sensing images. Different from the previous GAN-based super-resolution models in which their discriminator takes in a single image at one time, in our model, the discriminator is specifically designed to take in a pair of images: a generated image and its HR ground truth, to make better discrimination of the inputs. We further introduce a dual pathway network architecture, a random gate, and a coupled adversarial loss to learn the better correspondence between the discriminative results and the paired inputs. Experimental results on two public datasets demonstrate that our model can obtain more accurate super-resolution results in terms of both visual appearance and local details compared with other state-of-the-arts. Our code will be made publicly available.**

***Index Terms*—Super-resolution, remote sensing images, deep convolutional neural networks, generative adversarial networks (GAN), coupled adversarial training**

I. Introduction

Image super-resolution, as an important image processing technique that recovers high-resolution images from low-resolution ones, has been received great attention in recent years and has been widely used in many fields such as medical image enhancement [1] and small object detection [2]. In the field of remote sensing analysis, high-resolution images play an important role in many applications, including target detection [3–5], semantic labeling [7, 8, 14, 15], and scene analysis [9]. Apart from developing physical imaging technologies, image super-resolution provides an alternative way to effectively produce high-resolution remote sensing images [6, 10, 12, 16].

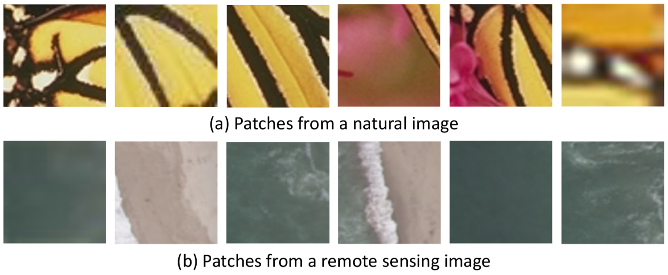
The essence of image super-resolution can be considered as the learning of an "universal" prior from image data, and then recovering the missing details of the low-resolution images by using this prior knowledge. In a remote sensing image super-resolution task, most of the previous methods are borrowed from the computer vision community without considering the nature of the remote sensing imaging process. Some early methods include sparse reconstruction based methods [6, 13] and discrete wavelet transform based methods [10]. These methods are designed by using low-level features and thus their performance is limited. In recent years, deep learning methods have played an important role in image super-resolution by constructing hierarchical convolutional architectures and learning high-level feature representations [16–18]. More recently, the residual dense model and recursive block are also introduced to further improve the performance of super-resolution [19–21]. Most of these methods simply focus on minimizing the mean squared reconstruction error and are evaluated by peak signal-to-noise ratios (PSNR). As a result, their super-resolved outputs may have very high PSNR, but may still suffer from an "over-smoothed" results [23–25]. To this end, the Generative Adversarial Networks (GAN) [26], has been introduced to image super-resolution more recently to perceptually improve the reconstruction results [22, 25, 29, 31].

GAN was originally proposed by A. Goodfellow *et al.* in 2014, and has then received great attention. GAN has achieved impressive results in various tasks such as image generation [26, 44], image style transfer [27, 28] and image superresolution [25, 33]. The key to GAN's success is the idea of an adversarial training framework under which the two networks, a generator G and a discriminator D, will contest with each other in a minimax two-player game and forces the generated data to be, in principle, indistinguishable from real ones. In a GAN-based image super-resolution task, a generator aims to generate high-resolution (HR) images from low-resolution (LR) ones while the discriminator aims to tell whether its input is real or not. Arguably, learning a discriminator D with a strong discriminating ability is essential for generating high-quality images. However, in remote sensing images, we discover through extensive statistical analysis that there are more flat regions and more low-frequency image components than the natural images, e.g. the areas of the desert and beach.

The work was supported by the National Key R&D Program of China under the Grant 2017YFC1405605, the National Natural Science Foundation of China under the Grant 61671037, the Beijing Natural Science Foundation under the Grant 4192034 and the National Defense Science and Technology Innovation Special Zone Project (*Corresponding author: Zhenwei Shi and Zhengxia Zou*).

Sen Lei and Zhenwei Shi (Corresponding author) are with Image Processing Center, School of Astronautics, Beihang University, Beijing 100191, China, and with Beijing Key Laboratory of Digital Media, Beihang University, Beijing 100191, China, and also with State Key Laboratory of Virtual Reality Technology and Systems, School of Astronautics, Beihang University, Beijing 100191, China (e-mail: senlei@buaa.edu.cn; shizhenwei@buaa.edu.cn).

Zhengxia Zou (Corresponding author) is with the Department of Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, MI 48109, U.S.A. (e-mail: zzhengxi@umich.edu).


(a) Patches from a natural image

(b) Patches from a remote sensing image

Fig. 1. To better understand the "discrimination-ambiguity" problem in a GAN-based super-resolution model, let's play a game. There are two groups of image patches, wherein each group, there is only one LR image patch and the rest are all HR ones. If you were a well-trained "discriminator", can you tell them apart? Obviously, one can easily identify that the last patch in the group (a) is an LR patch, but as for the group (b), things are not that easy.

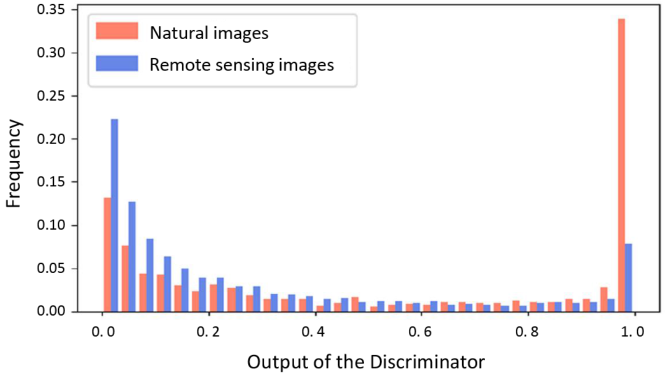


Fig. 2. The output distributions of a well-trained discriminator [25] on some natural image datasets: Set5 [34], Set14 [35], BSD100 [36] and Urban100 [37], and a remote sensing image datasets: UCMerced [45].

This may lead to a "discrimination-ambiguity" problem in which it will become more difficult for a discriminator to tell whether these image regions are generated or sampled from real HR images, even for human eyes. As a result, the quality of generative HR image may be deeply affected if we directly apply the above GAN-based super-resolution frameworks to remote sensing images.

To better understand this problem, let's play a game. In Fig. 1, we have two groups of image patches sampled from a natural image (a) and a remote sensing image (b). In each group, there is only one LR patch and the rest are HR ones. If you were a well-trained "discriminator", can you tell them apart? Obviously, one can easily identify that the last patch in the group (a) is an LR patch, but as for the group (b), things are not that easy. This is because even in an HR remote sensing image, it may still contain a large amount of low-frequency regions, which could bring unexpected bias to the adversarial training.

To further illustrate the "discrimination-ambiguity" problem, we run a well-trained discriminator [25] on two groups of datasets, a collection of natural image datasets ("Set5" [34], "Set14" [35], "BSD100" [36] and "Urban100" [37]), and a remote sensing image dataset "UCMerced" [45]. The output probabilities (how likely this region is identified that it belongs to a HR image) of the discriminator on a large set of random regions of each of two datasets are recorded. Fig. 2 shows their output distributions. We can see that for natural images, there are more output values close to probability 1, while for a remote sensing image, more outputs are close to probability 0. This indicates that remote sensing images contain more "ambiguity area" for a discriminator than natural ones.

Fig. 3 gives another example of the "discrimination-ambiguity" problem. We randomly select two images from the above two datasets, where the right half of each image has been down-sampled x4 and then up-sampled by using bicubic interpolation. The output scores of the discriminator on a random set of image regions are recorded. We can see that in the natural image (a), the HR part and LR part have a clear difference in their output values. However, in the remote sensing image (b), these image patches cannot be well distinguished, i.e. even in an HR part, the discriminator still gets relatively low outputs which are much close to that of an LR one.

To address the above problem, we propose a novel GAN-based super-resolution algorithm for remote sensing images, where we re-examine the discriminator under a totally different point of view, i.e. to formulate it as a coupled discriminative training process. We refer to our method as Coupled-Discriminated Generative Adversarial Networks (CDGAN). Different with all previous GAN based super-resolution models in which their discriminator takes in a single input image at one time, in our model, the discriminator is specifically designed to take in a pair of images: a generated image and its HR ground truth reference to make better discrimination of the inputs, especially for the low-frequency image regions. On this basis, we further introduce several technical components including a "dual pathway network architecture", a "random gate", and a "coupled adversarial loss function" to our model. Specifically, the dual pathway network and the random gate are designed to take in paired inputs for learning the better discrimination, while the coupled loss function is designed for learning better correspondence between the discriminative results and the paired inputs. We conduct our experiments on two publicly available datasets. Compared with other state of the art methods, our method obtains more accurate super-resolution results in terms of both visual appearance and local details.

The contributions of our work are summarized as follows:

- We propose a new GAN-based image super-resolution method named CDGAN for remote sensing images. In previous methods, the discriminator takes in a single input image at one time, while in our model, the discriminator is specifically designed to take in a pair of images to better discriminating its inputs especially for the low-frequency regions in a remote sensing image.
- We further introduce three additional technical components including a "dual pathway network architecture", a "random gate", and a "coupled loss function" in our discriminator in order to obtain better discrimination ability and learn the better correspondence between the outputs and the input pairs.

The rest parts of this paper are organized as follows. In Section II, we give a detailed description of the proposed

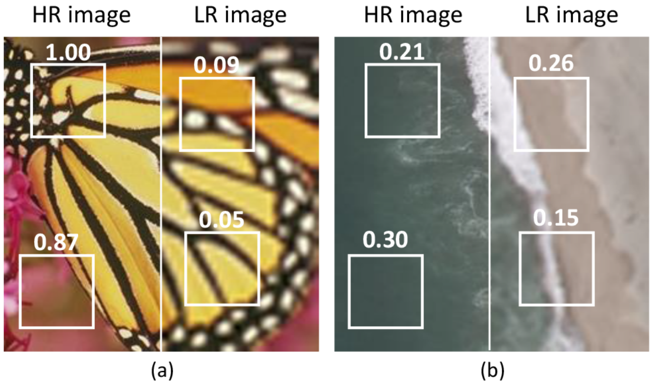


Fig. 3. Output probability (how likely a image patch belongs to a HR image) of a well-trained discriminator [25] on (a) a natural image, and (b) a remote sensing image.

method, including the network architectures, loss functions, and the implementation details. In Section III, we give a detailed description of our experimental datasets, evaluation metrics, ablation analysis and experimental results. The conclusions are drawn in Section IV.

II. Methodology

In this section, we will first give a brief introduction to the previous GAN-based image super-resolution methods. Then we will introduce the proposed method including the design of the network architectures, loss functions, and implementation details.

A. GAN-based image super-resolution

A typical GAN consists of two neural networks: a generator network G and a discriminator network D. The generator aims to learn a mapping $G(z)$ from a latent space $z \in \mathbb{Z}$ to a particular data distribution of interest. The discriminator, on one hand, aims to discriminate between instances from the true data distribution $x \sim p_{data}$ and those generated ones $G(z)$, on the other hand, feeds its output back to G to further make the generated data indistinguishable. The training of a GAN can be considered as as the following min-max problem:

$$\begin{aligned}\min_G \max_D \mathcal{L}(G,D) = \mathbb{E}_{z\sim p_z(z)}\{\log(1-D(G(z)))\} \\ + \mathbb{E}_{x\sim p_{data}}\{\log D(x)\},\end{aligned} \tag{1}$$

where x and z represent a true data point and an input random noise

In a GAN-based image super-resolution task [25, 33], the above random noise z will be replaced by a low-resolution image x. In addition, the generator G and the discriminator D are usually constructed based on deep convolutional networks. In this case, the G is trained to map a LR image x to a HR one, and the D is trained to distinguish a real HR image from a generated one. The G and D are trained to compete with each other. Their objective function can be rewritten as the follows:

$$\begin{aligned}\min_G \max_D \mathcal{L}(G,D) = \mathbb{E}_{x\sim p(x)}\{\log(1-D(G(x)))\} \\ + \mathbb{E}_{y\sim p(y)}\{\log D(y)\}\end{aligned} \tag{2}$$

where x and y represent a LR image and the corresponding HR ground truth. As the adversarial training progresses, the D will have more powerful discriminative ability and thus the HR images generated by G will become more and more realistic.

B. Coupled Discriminated GAN

To address the above mentioned “discrimination-ambiguity” problem in remote sensings, we propose a new GAN-based super-resolution method named Coupled Discriminated GAN. The flowchart of the proposed method is shown in Fig. 4. Our model consists of a generator G and a coupled discriminator D. Suppose X represents and LR image domain, Y represents an HR image domain. In this way, $x_i \in X$ represents an LR remote sensing image and $y_j \in Y$ represents its corresponding HR ground truth.

Generator: We design its architecture by referring to some recent image super-resolution methods [25, 32, 33] which achieve state-of-the-art performance in natural images. Instead of making any interpolation of the LR images before feeding it in the generator, we directly use it as the input to reduce additional computational overhead. A set of convolution and upsampling blocks are designed for the reconstruction. We further integrate the idea of residual learning [42] to ease the training. The ReLU activation function [38] is used after each layer of convolution only except for its output layer. We use the fractional-strided convolution [44] (a.k.a. the transposed convolution) for the upsampling of the feature maps. The details of the residual block and the upsample block are shown in the right part of Fig. 4 and the lower part of Table I. We use 16 residual blocks in the generator.

Coupled Discriminator: In this part, we will introduce the core of the proposed method: the coupled discriminator. The inputs of our discriminator consists of a super-resolved image produced by the generator and a HR ground truth reference. The coupled discriminator will be trained to distinguish which one is a real HR image and which one is generated. To learn better correspondence between the input pair and the discriminative outputs, the paired inputs are first fed into a random gate $\psi_z(t_1, t_2)$ that randomly shuffles the order of the two images and then produce a corresponding bool random variable d_z ($d_z = 0$ or 1) according to their input order, which is defined as follows:

$$(I_1, I_2, d_z) = \psi_z(t_1, t_2) = \begin{cases}(t_1, t_2, 1), & \text{if } z \geq 0.5 \\ (t_2, t_1, 0), & \text{if } z < 0.5\end{cases} \tag{3}$$

where z is uniformly distributed random variable in the range of [0, 1] to randomly adjust orders of the two input images I_1 and I_2. We set $t_1 = y$ and $t_2 = G(x)$, where $G(x)$ represents a super-resolved image and y represents a corresponding HR ground truth.

The two images are then fed into a dual-pathway network to extract features of the input image pairs. Then the feature maps of the two images are concatenated together and are further proceeded to produce the decision outputs. The lower-left part of Fig. 4 shows the structure of the dual-pathway network. As is suggested by [44], we use strided-convolutional layers instead of pooling layers, and use Leaky ReLU [39] activation

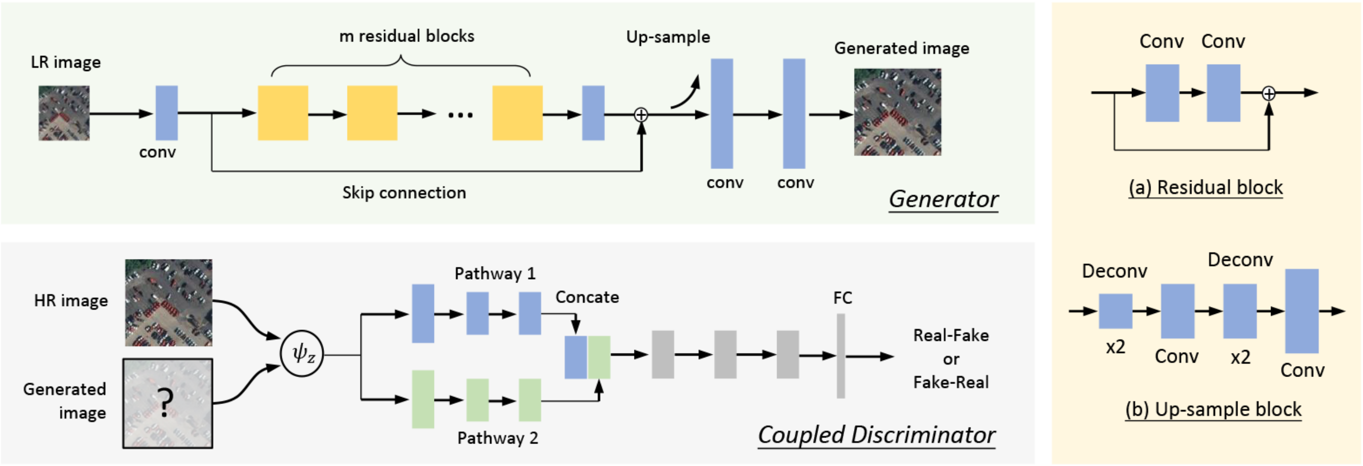


Fig. 4. An overview of the proposed method. The upper left part of this figure shows the architecture of our generator. The lower left part of this figure shows the architecture of our coupled discriminator. A detailed configuration of our model is shown in Table I.

function after all convolutional layers in our discriminator, except for the output layer, which uses sigmoid to convert the logits to probabilities.

C. Loss function

We frame the training of our coupled discriminator under a binary classification paradigm and we use d_z as its reference class id. When $I_1 = y$ and $I_2 = G(x)$, d_z is set to 1, and in this case the discriminator is trained to produce a high output probability that close to 1: $D(\psi_z(y, G(x))) \to 1$. In contrast, when $I_1 = G(x)$ and $I_2 = y$, d_z is set to 0, and in this case the discriminator is trained to produce a low output probability that close to 0: $D(\psi_z(y, G(x))) \to 0$.

The objective function of our coupled discriminator can be considered as the sum of the binary cross-entropy function, and thus can be written as follows:

$$\begin{aligned}\mathcal{L}_{adv}(D) = \mathbb{E}_{x,y\sim p_{data}}\{&d_z \log(D(\psi_z(y, G(x))))\\ &+ (1 - d_z)\log(1 - D(\psi_z(y, G(x))))\}.\end{aligned} \tag{4}$$

Our discriminator can be trained by maximizing the above objective function. As for the generator, it is trained to let D make more mistakes. Therefore, its objective function has a similar form with (4), while the only difference is to exchange their ground truth labels d_z and $(1 - d_z)$:

$$\begin{aligned}\mathcal{L}_{adv}(G) = -\mathbb{E}_{x,y\sim p_{data}}\{&d_z \log(1 - D(\psi_z(y, G(x))))\\ &+ (1 - d_z)\log(D(\psi_z(y, G(x))))\}.\end{aligned} \tag{5}$$

The adversarial training process of G and D can be essentially considered as a minimax optimization process, where G tries to minimize its objective while D tries to maximize it: $G^\star, D^\star = \arg\min_G \max_D \mathcal{L}_{adv}(G) + \mathcal{L}_{adv}(D)$.

TABLE I
A Detailed Configuration of our Model

	Layer names	Input	#Kernels	Sizes/Strides
Generator	C1	Image	64	$3\times3/1$
	m Res_blks	C1	–	–/–
	C2	Res_blk out	64	$3\times3/1$
	Upsample	C2+C1	–	–/–
	C3	Upsample	64	$3\times3/1$
	C4	C3	3	$3\times3/1$
Discriminator	P1/C1	Image	64	$3\times3/2$
	P2/C1	Image	64	$3\times3/2$
	P1/C2	P1/C1	128	$3\times3/2$
	P2/C2	P2/C1	128	$3\times3/2$
	P1/C3	P1/C2	256	$3\times3/2$
	P2/C3	P2/C2	256	$3\times3/2$
	C4	P1/C3+P2/C3	256	$3\times3/2$
	C5	C4	256	$3\times3/2$
	FC6	C5	–	–/–
Res_blk	C1	Input	64	$3\times3/1$
	C2	C1	64	$3\times3/1$
	Res_add	Input+C2	–	–/–
Upsample	DC1	Input	64	$3\times3/2$
	C1	DC1	64	$3\times3/1$
	DC2	C1	64	$3\times3/2$
	C2	DC2	64	$3\times3/1$

As is suggested by previous GAN-based super-resolution works [33, 58], in addition to the adversarial objective functions, we also introduce a "content loss" by minimizing the mean square error between a HR ground truth image and a generated one. This is because when we only use the adversarial loss for training, the model will usually introduce some undesired artifacts and make the generated image distorted since the generator simply ignores holding on the low-frequency contents (such as the structures and colors) while only focusing on high-frequency components. The content loss function is defined as follows:

$$\mathcal{L}_{content}(G) = \mathbb{E}_{x,y\sim p_{data}}\{\|G(x) - y\|_2^2\}. \tag{6}$$

We compute the mean square error $\|G(x) - y\|_2^2$ in their pixel space.

Our final objective function $\mathcal{L}(G, D)$ is defined as follows:

$$\mathcal{L}(G, D) = \mathcal{L}_{adv}(G) + \mathcal{L}_{adv}(D) + \lambda\mathcal{L}_{content}(G), \tag{7}$$

where $\lambda > 0$ controls the balance between the adversarial loss and the content loss. We aim to solve:

$$G^{\star}, D^{\star} = \arg\min_{G}\max_{D}\mathcal{L}(G, D). \tag{8}$$

When λ is set to a very large value, the proposed method will degenerate into a standard mean square error based super-resolution method.

D. Implementation Details

Our training samples are a set of 64×64 image patches that are randomly cropped from HR images and their corresponding LR ones. The training images are augmented by making random flip and rotation during training. We set $\lambda = 10^4$.

Table I gives a detailed configuration of the coupled discriminator and the generator, where "C" represents a convolution layer, "FC" represents a full-connected layer, "DC" represents a fractional-strided convolution layer, and "Pi" (i=1 or 2) represents the i'th pathway. The slope of Leaky ReLU activation in the coupled discriminator is set to 0.2.

As for the optimization, we train our model from scratch with Xavier initializer by alternatively updating D and G. we use the Adam optimizer [47] with the initial learn rate = 10^{-4}, the weight decay = 10^{-4}, and the mini-batch size = 16. Specifically, we first pre-train our generator only based on the content loss (6) for 10^5 iterations to provide a better initialization for the subsequent adversarial training, and then train our model based on the whole objective function (7) for 10^5 iterations where the learn rate decreases to half every 2.5×10^4 iterations. A complete training process of our method is summarized as follows:

- **Stage I.** Pre-train the generator G by minimizing the content loss $\mathcal{L}_{content}(G)$ for 10^5 iterations.
- **Stage II.** Alternatively update the discriminator D and the generator G based on (7) for 10^5 iterations:
  1) Fix G and update D by maximizing $\mathcal{L}(G, D)$.
  2) Fix D and update G by minimizing $\mathcal{L}(G, D)$.
  3) Repeat 1) and 2) until reach the max-iteration number.

III. Experimental Results and Analyses

In this section, we will give a detailed description of our experimental datasets, evaluation metrics, ablation analysis, and experimental results.

A. Experimental Dataset

We use two public datasets, "UCMerced" [45] and "WHU-RS19" [46], which have been commonly used in previous remote sensing image super-resolution literature [16, 18], to evaluate our method. For each dataset, we randomly select 40% images for training, 10% images for validation, and the rest for test. The original images are considered as HR images and we down-sample each image x4 as LR ones.

- **UCMerced dataset** [45]. This dataset consists of 21 classes of remote sensing scenes, including: agricultural, airplane, baseball-diamond, beach, etc. There are 100 images for each class. All images are in 256×256 pixels with a spatial resolution of 0.3m/pixel.

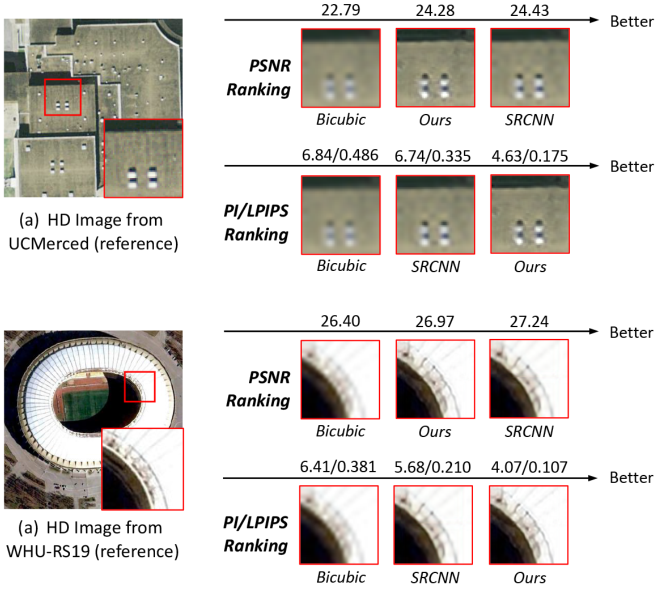


Fig. 5. An visual comparison of the three metrics: PSNR, PI [51], and LPIPS [56]. As suggested by some recent works [23–25], an algorithm may have high PSNR but suffer from an over-smoothed super-resolved output. In contrast, the evaluation ranking of PI and LPIPS are more consistent with their visual quality. (Zoom in for a better view.)

- **WHU-RS19 dataset** [46]. This dataset consists of 1,005 images in 19 classes of remote sensing scenes, including: airport, beach, bridge, commercial, etc. All images are in 600 × 600 pixels. The spatial resolution is up to 0.5m/pixel.

Furthermore, we test our method on some real-world multi-spectral images (3.2 m/pixel) from GaoFen-2 (GF-2) satellite. The three visible bands of these images are extracted and stacked into a pseudo-RGB for experiments. Since there are no corresponding high-resolution references, the super-resolved results are displayed and compared with other methods visually.

B. Evaluation Metrics

The PSNR has been widely used in image super-resolution community [32, 58] as a standard evaluation metric. Since the definition of PSNR is $10\log_{10}\frac{\max_I^2}{MSE}$, minimizing mean square loss is equal to maximizing PSNR. However, as suggested by some recent works [23–25], people notice that sometimes an algorithm may have high PSNR, but tends to get over-smoothed results and may also lack realistic visual appearance. To this end, apart from PSNR, we also use the Perception Index (PI) [51] and the Learned Perceptual Image Patch Similarity (LPIPS) [56] as two additional evaluation metrics in our experiments. The PSNR and LPIPS are implemented in Python and the PI is implemented in Matlab[1]. All the super-resolved results of the proposed method and comparing methods are evaluated by the same set of code.

[1] The implementation of the evaluation metrics can be found in the following websites. PSNR: https://github.com/scikit-image/scikit-image, PI: https://github.com/roimehrez/PIRM2018, LPIPS: https://github.com/richzhang/PerceptualSimilarity.

The PI was originally introduced as a no-reference image quality assessment method based on the low-level statistical features and is recently used in some super-resolution works [33, 51], by incorporating the criteria of “MA” [53] and “NIQE” [54]:

$$\mathrm{PI} = \frac{1}{2}((10 - \mathrm{MA}) + \mathrm{NIQE}), \tag{9}$$

In an image super-resolution task, a lower PI indicates a better super-resolved result.

Since PI is a no-reference measurement, we further use a full-reference metric named Learned Perceptual Image Patch Similarity (LPIPS) [56] as an alternative evaluation metric. The LPIPS measures perceptual image similarity using a pre-trained deep network, which can be computed as the l_2 distance between a super-resolved images $G(x)$ and an HR reference image y in their feature space:

$$\mathrm{LPIPS}(y, \hat{y}) = \sum_{l} \frac{1}{N_l} \|\omega_l \odot (\phi(y)_l - \phi(\hat{y})_l)\|_2^2, \tag{10}$$

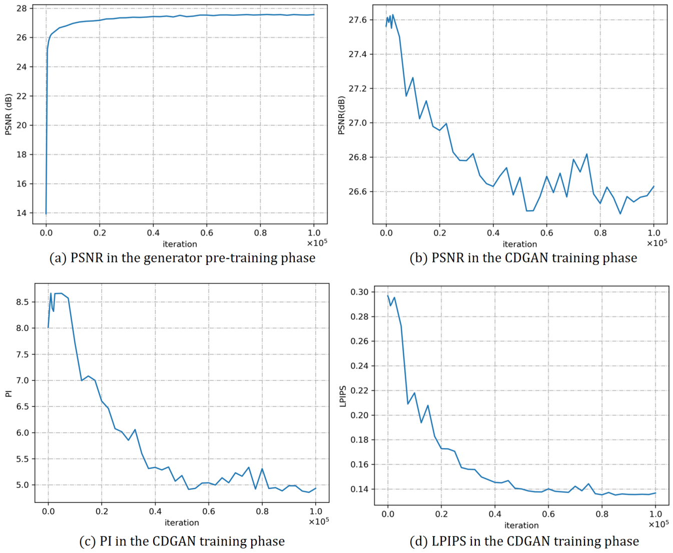
where $\phi(\cdot)_l$ represents a feature space constructed by a well-trained deep CNN of its l’th layer, and N_l is the number of elements in $\phi(\cdot)_l$. ω_l is a learned weight vector and $\odot$ is the channel-wise product operation. A lower LPIPS indicates a better super-resolved result.

Fig. 5 gives an visual comparison of the three metrics. The two example images are from UCMerced dataset and WHU-RS19 dataset. The output of bicubic interpolation, SRCNN [58], and our method on the three different metrics are recorded. We can observe the results of SRCNN have a higher PSNR than our CDGAN but they suffer from a “blurring effect”. In contrast, the evaluation ranking of PI and LPIPS are more consistent with their visual quality. Nevertheless, in our following experiments, we still use PSNR as a reference.

In Fig. 6, we show the evolution of these three metrics on the UCMerced validation dataset. For PSNR, larger values indicate better. For PI and LPIPS, smaller values indicate better. Fig. 6 (a) shows the PSNR curve in the pre-training phase of the generator, and Fig. 6 (b)-(d) show the PSNR, PI and LPIPS curves in the alternative training phase of the generator and the coupled discriminator. It should be noticed that although the PSNR decreases along with the alternative training, the perceptual quality of super-resolved images keeps improving, which is suggested by the decrease of the PI and LPIPS.

C. Ablation Studies

The ablation studies are conducted to analyze the importance of each component of the proposed method, including the coupled adversarial loss (Cp-Adv-Loss) and the content loss (Cont-Loss), as is shown in Table II. We also compare with a baseline GAN-based method that is only trained with the standard adversarial loss (Std-Adv-Loss). A weak baseline method, bicubic interpolation, is first evaluated, then we gradually integrate these techniques. All evaluations of the ablation analyses are performed based on the same set of configurations.



(a) PSNR in the generator pre-training phase

(b) PSNR in the CDGAN training phase

(c) PI in the CDGAN training phase

(d) LPIPS in the CDGAN training phase

Fig. 6. The metric curves including PSNR, PI and LPIPS in the training phase.

TABLE II
RESULTS OF THE ABLATION ANALYSIS ON: COUPLED ADVERSARIAL LOSS (CP-ADV-LOSS), CONTENT LOSS (CONT-LOSS), AND STANDARD ADVERSARIAL LOSS (STD-ADV-LOSS). ALL METHODS ARE TRAINED AND EVALUATED ON UCMERCED DATASET.

Cp-Adv-Loss	Cont-Loss	Std-Adv-Loss	PSNR / PI / LPIPS
×	×	×	25.34 / 7.483 / 0.464
✓	×	×	16.21 / 8.918 / 0.558
×	✓	×	**27.56** / 8.012 / 0.297
×	✓	✓	26.07 / **4.855** / 0.195
✓	✓	×	26.63 / 4.933 / **0.137**

- **Cp-Adv-Loss**: we only train our generator based on the coupled adversarial loss $\mathcal{L}_{adv}(G) + \mathcal{L}_{adv}(D)$ without any help of content loss.
- **Cont-Loss**: we only train our generator based on the image content loss $\mathcal{L}_{content}(G)$ without any help of adversarial training, which is similar to the SRCNN [58].
- **Cont-Loss + Std-Adv-Loss**: we train our generator based on content loss $\mathcal{L}_{content}(G)$ and a standard adversarial loss, which is similar to the SRGAN [25].
- **Cont-Loss + Cp-Adv-Loss**: we train our generator based on content loss $\mathcal{L}_{content}(G)$ and our proposed coupled adversarial loss $\mathcal{L}_{adv}(G) + \mathcal{L}_{adv}(D)$.

Table II shows their evaluation accuracy. For each evaluation metric, the best result is marked as bold and the second best is marked with an underline. As we can see, the integration of the “adversarial training” and “content loss” yields noticeable improvements of the reconstruction accuracy on LPIPS and PI. Particularly, when we apply the proposed “coupled adversarial loss”, we obtain the best result on LPIPS. Fig. 7 shows two examples with the different combinations of loss functions. Our method obtains more accurate super-resolution results in terms of both visual appearance and local details.

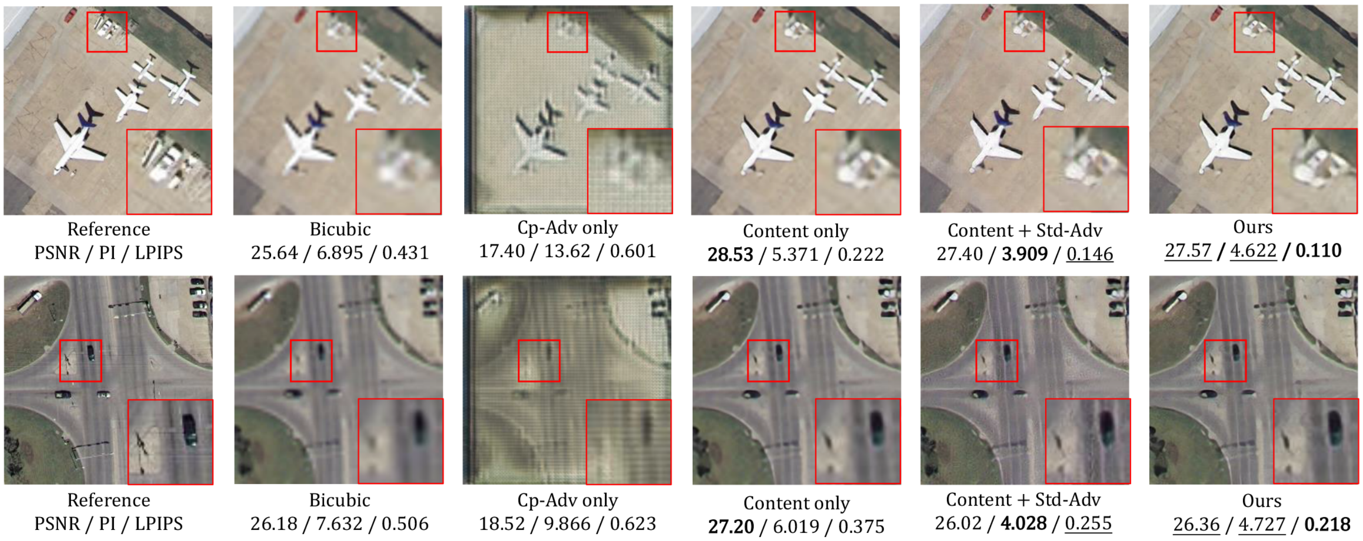


Fig. 7. Some super-resolved outputs of our ablation experiments. Image from the first row and the second row are from UCMerced dataset and WHU-RS19 dataset respectively. (Zoom in for a better view.)

D. Comparison with Other Methods

We further compare our method with some state of the super-resolution methods including SRCNN [58], LGCNet [16], EDSR [59], RCAN [60] and SRGAN [25] on UCMerced dataset and WHU-RS19 dataset. Among these methods, SRCNN, LGCNet, EDSR and RCAN are recently proposed CNN-based methods that minimize the content loss (L2 loss or L1 loss). SRGAN is a more recent GAN-based method that integrates content loss and the standard adversarial loss. All these methods are fully optimized on our train + validation data to obtain their best performance for a fair comparison.

Table III and Table IV shows the comparison results of different methods on UCMerced dataset and WHU-RS19 dataset respectively. Our method achieves the best LPIPS performance on both datasets, with the lower PI and LPIPS comparing with the CNN-based methods and with the higher PSNR and the lower LPIPS comparing with SRGAN. Fig. 8 shows some super-resolved examples, where the images of the rows (a)-(b) are from UCMerced dataset and the images of the rows (c)-(d) are from WHU-RS19 dataset. Comparing with other methods, the proposed method obtains better perceptual performance with more details and textures. Although SRGAN achieves the lowest PI, its super-resolved results are often affected by undesired artifacts.

In Fig. 9, we show some super-resolution reconstruction results of the GF-2 satellite data. Here, we use two non-reference image quality metrics, i.e., NIQE[54] and SSEQ[55], to assess these super-resolved results. For either of the two indicators, a lower score indicates a better reconstruction result. Compared with some CNN-based methods, e.g., SRCNN, LGCNet, EDSR and RCAN, the proposed CDGAN obtains lower NIQE and SSEQ with more clear edges. It also should be noted that although SRGAN achieves the lowest NIQE and SSEQ, it suffers from some checkerboard artifacts.

In Table V, we report three different indices on model efficiency, i.e., the number of model parameters, the number of floating-point operations (FLOPs)[2] and the inference time (in GPU and CPU mode). We use the WHU-RS19 dataset to compute FLOPs and inference time, where the sizes of LR images and HR ground truth reference images are 125 × 125 and 600 × 600 pixels respectively. The GPU runtime is tested with an Nvidia GeForce GTX 1080Ti graphics card and the CPU runtime is tested with an Intel i7-6700K CPU and 32GB RAM. Comparing with other methods, the runtime of the proposed method in GPU mode is the smallest one and the runtime in CPU mode is only larger than SRCNN. It should be noted that since the inference time for GAN-based methods is determined by their generator, we here only list the parameters of the generator in Table I. As shown in Table V, although the proposed CDGAN has 20 times parameters than the SRCNN, it has less inference time. We believe the reason behind this phenomenon is the different input size of the two networks. In our experiments. The proposed CDGAN uses a 125 × 125 LR image as its input. However, the SRCNN uses a 600 × 600 image as its input by firstly upsampling the 125 × 125 LR image to the same size of a HR one before feeding it into the network. The larger input size will inevitably cause a heavier computational overhead for the SRCNN although it has less parameters.

IV. Conclusion

We proposed a new GAN-based super-resolution method named Coupled Discriminated GAN for remotes sensing images. Previous GAN-based image super-resolution methods suffer from a “discrimination-ambiguity” when dealing with the low-frequency image regions, and thus the quality of the super-resolved images could be deeply affected. Different from a previous GAN-based method where the discriminator only takes in a single image at one time to classified whether it is a real HR image or a generated one, our discriminator takes in

[2] The model parameters and FLOPs can be counted by the following repository: https://github.com/Lyken17/pytorch-OpCounter

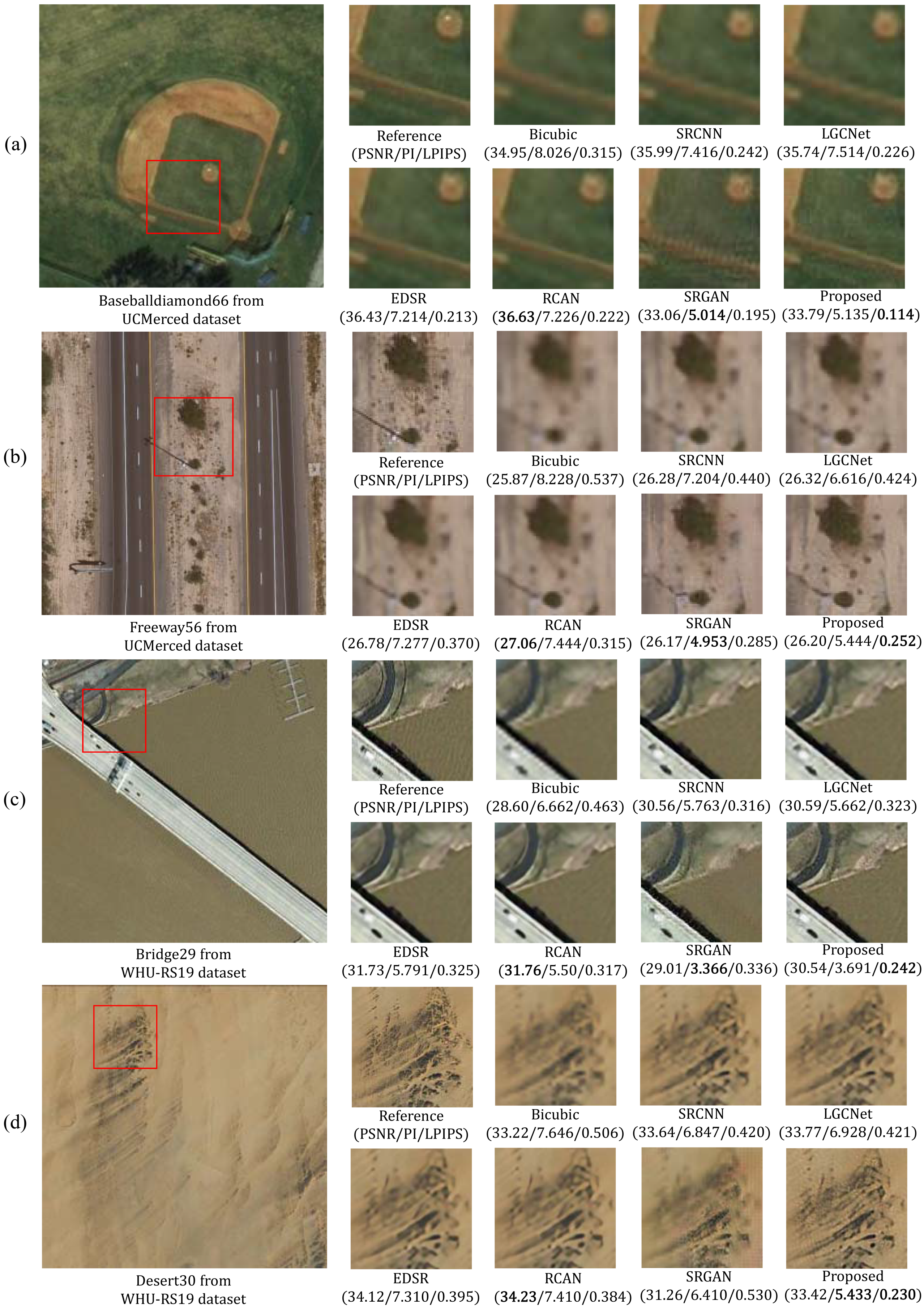


Fig. 8. A Comparison of the super-resolved outputs with different methods: Bicubic, SRCNN [58], LGCNet [16], EDSR [59], RCAN [60], SRGAN [25], and the proposed method. (Zoom in for a better view.)

TABLE III
A COMPARISON OF DIFFERENT METHODS ON THE UCMERCED TEST SET. FOR PSNR, A HIGHER SCORE INDICATES BETTER. FOR PI AND LPIPS, A LOWER SCORE INDICATES BETTER.

class	Bicubic PSNR/PI/LPIPS	SRCNN [58] PSNR/PI/LPIPS	LGCNet [16] PSNR/PI/LPIPS	EDSR [59] PSNR/PI/LPIPS	RCAN [60] PSNR/PI/LPIPS	SRGAN [25] PSNR/PI/LPIPS	CDGAN (ours) PSNR/PI/LPIPS
agricultural	25.50/11.38/0.618	25.87/12.03/0.566	25.94/11.21/0.556	26.05/12.01/0.531	26.08/12.83/0.519	25.13/10.84/0.479	25.00/11.45/0.493
airplane	24.85/7.096/0.448	26.34/6.546/0.314	26.72/6.311/0.299	27.39/6.120/0.249	27.76/5.940/0.236	25.98/3.668/0.172	26.42/4.478/0.140
baseballdiamond	31.40/7.749/0.409	32.49/7.098/0.332	32.53/6.917/0.308	32.88/7.003/0.300	33.09/7.003/0.295	30.41/5.159/0.203	31.02/5.101/0.172
beach	34.22/8.281/0.335	34.90/7.831/0.269	35.07/7.295/0.244	35.30/7.856/0.240	35.55/7.804/0.234	32.43/6.692/0.198	32.94/6.076/0.155
buildings	23.12/6.935/0.460	24.90/6.730/0.299	25.28/6.506/0.277	26.10/6.622/0.220	26.41/6.084/0.197	24.44/3.732/0.169	24.66/4.840/0.153
chaparral	23.45/8.349/0.575	24.35/9.859/0.448	24.48/10.08/0.430	24.81/15.33/0.435	24.90/16.50/0.427	22.95/6.085/0.198	23.69/11.58/0.220
denseresidential	23.74/7.071/0.488	25.31/7.725/0.324	25.67/7.449/0.300	26.35/7.438/0.251	26.78/7.254/0.228	24.70/3.730/0.179	25.08/4.927/0.161
forest	25.90/7.576/0.615	26.43/7.611/0.537	26.48/7.551/0.514	26.60/8.105/0.537	26.65/8.623/0.518	24.95/3.534/0.271	25.46/4.632/0.260
freeway	25.66/8.583/0.437	26.97/8.862/0.321	27.32/8.478/0.303	28.04/9.079/0.252	28.48/9.057/0.232	26.56/5.659/0.185	26.67/6.055/0.170
golfcourse	33.28/8.058/0.368	34.48/7.337/0.291	34.62/7.174/0.271	34.78/7.399/0.269	34.98/7.323/0.272	31.31/5.560/0.174	32.61/5.454/0.120
harbor	19.41/7.373/0.444	20.62/6.899/0.297	20.88/6.845/0.273	21.72/7.074/0.215	22.10/6.502/0.180	20.77/5.436/0.199	20.40/5.317/0.148
intersection	24.62/7.139/0.479	25.81/7.172/0.347	26.02/6.974/0.327	26.53/6.912/0.282	26.91/6.774/0.242	25.06/3.528/0.198	25.18/4.472/0.194
mediumresidential	23.73/7.003/0.524	25.05/7.228/0.366	25.32/7.053/0.341	25.92/6.910/0.305	26.32/6.731/0.280	24.22/3.136/0.200	24.68/4.473/0.189
mobilehomepark	20.38/7.210/0.539	21.93/8.267/0.349	22.17/8.101/0.330	22.96/7.145/0.268	23.43/7.042/0.234	21.68/3.993/0.207	21.67/5.021/0.205
overpass	23.27/8.079/0.520	24.21/8.304/0.405	24.61/7.915/0.376	25.46/7.581/0.298	24.99/6.686/0.264	24.29/4.345/0.212	24.41/5.005/0.215
parkinglot	19.73/7.164/0.441	20.90/6.994/0.307	21.08/6.753/0.293	21.54/6.542/0.246	22.08/5.598/0.177	20.74/3.853/0.178	20.41/5.166/0.162
river	26.60/7.544/0.532	27.30/6.937/0.455	27.39/6.506/0.429	27.47/7.015/0.424	27.49/6.811/0.412	25.92/4.060/0.269	26.25/4.314/0.251
runway	26.62/8.434/0.420	27.82/8.091/0.319	28.46/7.597/0.297	29.37/9.031/0.247	29.42/8.672/0.245	27.91/6.519/0.184	28.14/6.632/0.158
sparseresidential	27.13/7.155/0.445	28.39/6.643/0.328	28.58/6.645/0.311	29.05/6.767/0.303	29.26/6.561/0.300	26.88/3.369/0.185	27.53/4.213/0.163
storagetanks	28.06/7.311/0.432	29.44/6.742/0.312	29.76/6.575/0.289	30.39/6.385/0.241	30.46/6.159/0.229	28.42/4.455/0.187	28.55/4.667/0.162
tenniscourt	27.58/7.271/0.423	28.81/6.921/0.293	28.99/6.771/0.277	29.65/6.688/0.227	30.01/6.663/0.214	27.58/3.607/0.183	27.97/4.338/0.162
average	25.63/7.751/0.474	26.78/7.706/0.356	27.02/7.462/0.336	27.54/7.857/0.302	**27.77**/7.743/0.283	25.82/**4.807**/0.211	26.13/5.629/**0.193**

TABLE IV
A COMPARISON OF DIFFERENT METHODS ON THE WHU-RS19 TEST SET. FOR PSNR, A HIGHER SCORE INDICATES BETTER. FOR PI AND LPIPS, A LOWER SCORE INDICATES BETTER.

class	Bicubic PSNR/PI/LPIPS	SRCNN [58] PSNR/PI/LPIPS	LGCNet [16] PSNR/PI/LPIPS	EDSR [59] PSNR/PI/LPIPS	RCAN [60] PSNR/PI/LPIPS	SRGAN [25] PSNR/PI/LPIPS	CDGAN (ours) PSNR/PI/LPIPS
airport	26.12/6.709/0.438	27.29/5.901/0.296	27.44/5.939/0.293	28.09/5.812/0.267	28.22/5.577/0.257	32.89/3.113/0.214	27.31/3.755/0.170
beach	42.41/8.399/0.125	40.62/7.038/0.141	42.51/7.560/0.086	42.16/7.765/0.086	43.58/7.961/0.093	32.86/7.369/0.120	40.93/6.351/0.054
bridge	31.48/7.288/0.290	32.90/6.257/0.210	33.17/6.328/0.205	34.26/6.431/0.197	34.44/6.434/0.198	21.73/4.523/0.205	32.91/4.673/0.135
commercial	22.04/7.066/0.552	23.20/6.451/0.391	23.29/6.472/0.382	23.78/6.583/0.370	23.90/6.434/0.350	26.16/3.346/0.239	23.13/4.848/0.259
desert	38.32/8.242/0.359	37.80/7.060/0.297	38.73/7.247/0.279	39.09/7.794/0.277	39.40/7.790/0.274	26.36/7.435/0.414	37.63/6.024/0.165
farmland	33.86/7.598/0.374	34.56/7.007/0.288	34.71/7.014/0.284	35.33/7.185/0.267	35.49/7.157/0.253	35.62/5.435/0.258	33.51/4.777/0.190
footballfield	25.69/6.624/0.418	27.05/5.841/0.286	27.24/5.920/0.283	28.03/5.793/0.257	28.19/5.633/0.240	31.29/2.868/0.204	27.17/3.644/0.169
forest	25.72/7.362/0.565	26.28/7.202/0.476	26.30/7.201/0.477	26.45/7.528/0.487	26.47/7.340/0.474	22.45/3.214/0.339	25.49/4.762/0.282
industrial	24.44/6.982/0.476	25.68/6.504/0.328	25.83/6.515/0.321	26.58/6.573/0.297	26.73/6.419/0.282	26.11/3.258/0.216	25.68/4.633/0.199
meadow	34.53/7.948/0.417	34.91/7.255/0.341	35.06/7.284/0.341	35.34/7.622/0.355	35.41/7.547/0.353	25.06/6.301/0.308	33.25/4.643/0.193
mountain	22.15/7.316/0.682	22.72/7.262/0.585	22.75/7.221/0.575	22.89/7.431/0.591	22.91/7.156/0.568	24.69/2.885/0.341	22.42/4.888/0.390
park	26.50/6.880/0.519	27.38/5.978/0.400	27.46/5.997/0.390	27.92/6.090/0.389	28.02/5.942/0.380	32.78/2.741/0.266	26.97/3.685/0.235
parking	24.66/6.910/0.369	26.10/6.380/0.228	26.21/6.386/0.228	27.18/6.526/0.189	27.54/6.542/0.181	25.19/3.557/0.196	25.99/4.553/0.133
pond	29.65/6.823/0.355	30.55/5.646/0.272	30.62/5.693/0.268	31.03/5.747/0.268	31.09/5.819/0.270	28.82/3.541/0.210	29.81/3.507/0.155
port	24.03/6.689/0.416	25.27/5.871/0.286	25.37/5.931/0.281	26.01/5.839/0.267	26.11/5.907/0.256	24.49/3.233/0.201	25.17/4.008/0.182
railwaystation	23.61/6.959/0.496	24.88/6.217/0.347	25.05/6.280/0.341	25.62/6.314/0.316	25.77/5.887/0.294	24.13/3.489/0.226	25.06/4.713/0.201
residential	21.19/7.384/0.555	22.55/7.388/0.383	22.69/7.372/0.373	23.36/7.221/0.360	23.53/7.137/0.340	21.96/3.696/0.210	22.74/5.518/0.253
river	26.75/6.730/0.496	27.56/5.775/0.386	27.61/5.849/0.383	27.94/5.860/0.394	27.99/5.885/0.393	26.11/2.897/0.258	26.95/3.569/0.232
viaduct	23.85/6.916/0.473	25.28/6.462/0.315	25.44/6.502/0.310	26.30/6.319/0.277	26.53/5.876/0.257	24.45/2.983/0.201	25.57/4.400/0.180
average	27.74/7.201/0.441	28.56/6.500/0.329	28.81/6.564/0.321	29.33/6.654/0.311	**29.54**/6.550/0.301	27.01/**3.994**/0.244	28.30/4.576/**0.199**

paired inputs to learn better discriminative ability for remote sensing images, especially for those low-frequency regions. We further design a dual pathway network architecture, a random gate, and a coupled adversarial loss in order to learn the better correspondence between the discriminative outputs and the paired inputs. We conduct our experiments on two public datasets and GF-2 satellite data. The ablation analysis suggests the effectiveness of the proposed couple adversarial training framework. Our method achieves better results than other state of the art methods in terms of both evaluation metrics and visual quality.

REFERENCES

[1] H. Greenspan, "Super-resolution in medical imaging," *Comput. J.*, vol. 52, no. 1, pp. 43-63, Jan. 2009.

[2] Y. Bai, Y. Zhang, M. Ding, *et al.* "SOD-MTGAN: Small Object Detection via Multi-Task Generative Adversarial Network," *European Conference on Computer Vision*, 2018: 206-221.

[3] A. J. Tatem, H. G. Lewis, P. M. Atkinson, and M. S. Nixon, "Super-resolution target identification from remotely sensed images using a Hopfield neural network,"

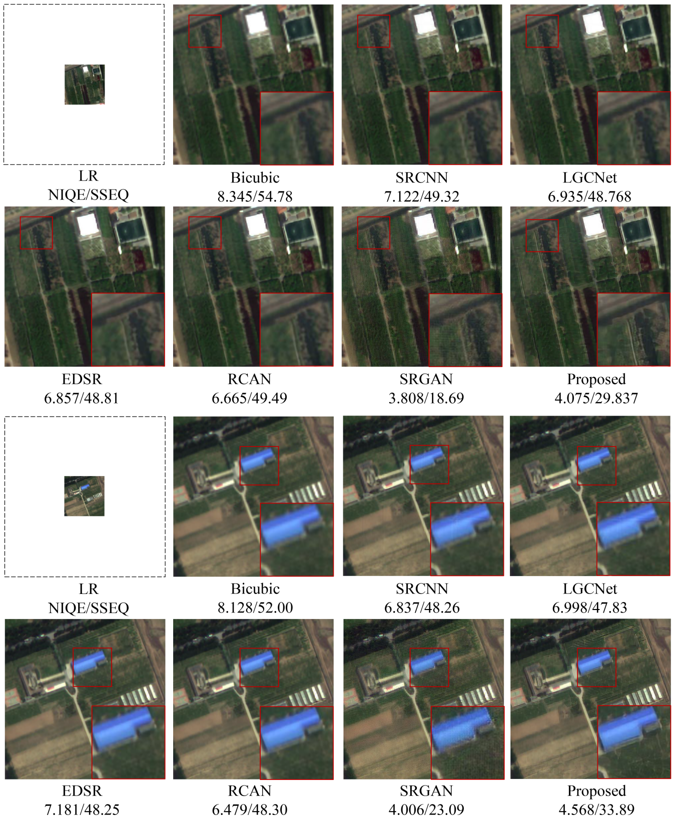


Fig. 9. The super-resolved results of GF-2 satellite data with different methods: Bicubic, SRCNN [58], LGCNet [16], EDSR [59], RCAN [60], SRGAN [25], and the proposed method. (Zoom in for a better view.)

TABLE V
COMPARISONS OF MODEL PARAMETERS, FLPOS AND GPU/CPU RUNTIME

Model	Params	FLOPs	GPU Runtime	CPU Runtime
SRCNN	69K	26.0G	0.065s	1.782s
LGCNet	193K	69.7G	0.076s	3.328s
EDSR	43M	1130G	0.251s	69.31s
RCAN	16M	359G	1.113s	51.08s
SRGAN	1.5M	50.2G	0.062s	4.913s
CDGAN	1.4M	74.5G	0.046s	3.154s

IEEE Trans. Geosci. Remote Sens., vol. 39, no. 4, pp. 781-796, Apr. 2001.

[4] Z. Zou and Z. Shi, "Ship Detection in Spaceborne Optical Image with SVD Networks," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 10, pp. 5832-5845, Oct. 2016.

[5] Z. Zou and Z. Shi, "Random access memories: A new paradigm for target detection in high resolution aerial remote sensing images," *IEEE Trans. on Image Process.*, vol 27, no. 3, pp. 1100-1111, Mar. 2018.

[6] Z. Pan, J. Yu, H. Huang, *et al.*, "Super-resolution based on compressive sensing and structural self-similarity for remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 9, pp. 4864-4876, 2013.

[7] H. Lin, Z. Shi and Z. Z, " Fully Convolutional Network with Task Partitioning for Inshore Ship Detection in Optical Remote Sensing Images," *IEEE Geoscience and Remote Sensing Letters*, vol 14, no. 10, pp. 1665-1669. Oct. 2017.

[8] X. Wu and Z. Shi, "Utilizing Multilevel Features for Cloud Detection on Satellite Imagery," *Remote Sensing*, 2018, 10(10): 1853.

[9] Z. Shi and Z. Zou, "Can a Machine Generate Human-like Language Descriptions for a Remote Sensing Image?," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 6, pp. 3623-3634, Jun. 2017.

[10] H. Chavez-Roman and V. Ponomaryov, "Super resolution image generation using wavelet domain interpolation with edge extraction via a sparse representation," *IEEE Trans. Geosci. Remote Sens.*, vol. 11, no. 10, pp. 1777-1781, 2014.

[11] J. Li, Q. Yuan, H. Shen, *et al.* "Hyperspectral Image Super-Resolution by Spectral Mixture Analysis and Spatial-Spectral Group Sparsity," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 9, pp. 1250-1254, 2016.

[12] L. Yue, H. Shen, J. Li, Q. Yuanc, H. Zhang, and L. Zhang, "Image superresolution: The techniques, applications, and future," *Signal Process.* vol. 128, pp. 389-408, Nov. 2016.

[13] B. Hou, K. Zhou, and L. Jiao, "Adaptive Super-Resolution for Remote Sensing Images Based on Sparse Representation With Global Joint Dictionary Model," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 4, pp. 2312-2327, 2018.

[14] B. Pan, Z. Shi, X. Xu, *et al.* "CoinNet: Copy Initialization Network for Multispectral Imagery Semantic Segmentation," *IEEE Geosci. Remote Sens. Lett.*, 2018.

[15] B. Pan, Z. Shi, X. Xu, "Analysis for the Weakly Pareto Optimum in Multiobjective-Based Hyperspectral Band Selection," *IEEE Trans. Geosci. Remote Sens.*, 2019.

[16] S. Lei, Z. Shi, and Z. Zou, "Super-Resolution for Remote Sensing Images via Local-Global-Combined Network," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 8, pp. 1243-1247, Aug. 2017.

[17] T. Wang, W. Sun, H. Qi, and P. Ren "Aerial Image Super Resolution via Wavelet Multiscale Convolutional Neural Networks," *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 5, pp. 769-773, 2018.

[18] Haut J M, Fernandez-Beltran R, Paoletti M E, *et al.* "A new deep generative network for unsupervised remote sensing single-image super-resolution, " *IEEE Geosci. Remote Sens. Lett.*, vol. 56, no. 11, pp. 6792-6810, 2018.

[19] Z. Pan, W. Ma, J. Guo, *et al.* "Super-Resolution of Single Remote Sensing Image Based on Residual Dense Backprojection Networks, " *IEEE Trans. Geosci. Remote Sens.*, 2019.

[20] Haut J M, Fernandez-Beltran R, Paoletti M E, *et al.* "Remote Sensing Image Superresolution Using Deep Residual Channel Attention, " *IEEE Trans. Geosci. Remote Sens.*, 2019.

[21] K. Jiang, Z. Wang, P. Yi, *et al.* "Deep distillation recursive network for remote sensing imagery super-resolution, " *Remote Sensing*, 2018, 10(11): 1700.

[22] K. Jiang, Z. Wang, P. Yi, *et al.* "Edge-Enhanced GAN for Remote Sensing Image Superresolution, " *IEEE Trans. Geosci. Remote Sens.*, 2019.

[23] Mathieu M, Couprie C, LeCun Y, "Deep multi-scale video prediction beyond mean square error," arXiv preprint arXiv:1511.05440, 2015.

[24] Johnson J, Alahi A, Fei-Fei L, "Perceptual losses for real-time style transfer and super-resolution," in *European Conference on Computer Vision*, 2016: 694-711.

[25] C. Ledig, L. Theis, F. Huszr, *et al*, "Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, 2017.

[26] I. Goodfellow, J. Pouget-Abadie, M. Mirza, *et al*, "Generative adversarial nets," in *Advances in neural information processing systems*, 2014.

[27] P. Isola, J. Zhu, T. Zhou, *et al*, "Image-to-image translation with conditional adversarial networks," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, 2017.

[28] J. Zhu, T. Park, P. Isola, *et al*, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Proc. IEEE International Conference on Computer Vision*, 2017: 2223-2232.

[29] X. Wang, K. Yu, S. Wu, *et al*, "Recovering realistic texture in image super-resolution by deep spatial feature transform," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, 2018: 606-615.

[30] L. Galteri, L. Seidenari, M. Bertini, *et al*, "Deep generative adversarial compression artifact removal," *arXiv preprint arXiv:1704.02518*, 2017.

[31] X. Xu, D. Sun, J. Pan, *et al*, "Learning to super-resolve blurry face and text images," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, 2017.

[32] B. Lim, S. Son, H. Kim, *et al*, "Enhanced deep residual networks for single image super-resolution," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, 2017: 136-144.

[33] X. Wang, K. Yu, S. Wu, *et al*, "Esrgan: Enhanced super-resolution generative adversarial networks," in *European Conference on Computer Vision*, 2018: 63-79.

[34] Bevilacqua, M., Roumy, A., Guillemot, C., Morel, M.L.A.: Low-complexity singleimage super-resolution based on nonnegative neighbor embedding. In: BMVC (2012)

[35] Zeyde, R., Elad, M., Protter, M., "On single image scale-up using sparsere presentations," in *Proceedings of 7th International Conference Curves Surfaces*, 2010

[36] Martin, D., Fowlkes, C., Tal, D., Malik, J., "A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics," in *Proc. IEEE Conf. Comput. Vis.*, 2001

[37] Huang, J.B., Singh, A., Ahuja, N., "Single image super-resolution from transformed self-exemplars," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2015.

[38] V. Nair, G. E. Hinton, "Rectified linear units improve restricted boltzmann machines," in *Proc. International Conference on Machine Learning* , 2010.

[39] Maas, Andrew L, Hannun, *et al*, "Rectifier nonlinearities improve neural network acoustic models," in *Proc. International Conference on Machine Learning*, 2013.

[40] Xu B, Wang N, Chen T, *et al*, "Empirical evaluation of rectified activations in convolutional network," arXiv preprint arXiv:1505.00853, 2015.

[41] J. Kim, J. K. Lee, and K. M. Lee, "Accurate image super-resolution using very deep convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 1646-1654.

[42] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 770-778.

[43] T. Hui, C. C. Loy, and X. Tang, "Depth map super-resolution by deep multi-scale guidance," in *Proc. Eur. Conf. Comput. Vis.* in 2016, pp. 353-369.

[44] Radford A, Metz L, Chintala S, "Unsupervised representation learning with deep convolutional generative adversarial networks," *arXiv preprint* arXiv:1511.06434, 2015.

[45] Y. Yang and S. Newsam, "Bag-of-visual-words and spatial extensions for land-use classification," in *Proc. ACM GIS*, 2010, pp. 270-279.

[46] G. Xia, W. Yang, J. Delon, *et al.*, "Structural high-resolution satellite image indexing," *ISPRS TC VII Symposium-100 Years ISPRS*, vol. 38, 2010.

[47] D. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization," in *International Conference on Learning Representations*, 2014.

[48] Abadi M, Barham P, Chen J, *etal.*, "Tensorflow: a system for large-scale machine learning," OSDI. 2016, 16: 265-283.

[49] Xu X, Sun D, Pan J, *et al.*, "Learning to Super-Resolve Blurry Face and Text Images," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017: 251-260.

[50] R. Yeh, C. Chen, T. Y. Lim, *etal.*, "Semantic Image Inpainting with Perceptual and Contextual Losses," *arXiv preprint* arXiv:1607.07539, 2016.

[51] Blau, Y., Mechrez, R., Timofte, R., *et al.*, "The 2018 PIRM Challenge on Perceptual Image Super-resolution ," *European Conference on Computer Vision.*, Springer, Cham, 2018: 334-355.

[52] Blau Y., Michaeli T., "The perception-distortion trade-off," *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018: 6228-6237.

[53] Ma, C., Yang, C.Y., Yang, X., *et al.*, "Learning a no-reference quality metric for single-image super-resolution," *Computer Vision and Image Understanding*, 2017: 1–16.

[54] Mittal, A., Soundararajan, R., Bovik, A.C., "Making a "completely blind" image quality analyzer," *IEEE Signal Process. Lett.*, vol. 20, no. 3, pp. 209-212, 2013.

[55] Liu L., Liu B., Huang H., *et al.*, "Making a "No-reference image quality assessment based on spatial and spectral entropies," *Signal Processing: Image Communication*, vol. 29, no. 8, pp. 856-863, 2014.

[56] R. Zhang, P. Isola, A. Efros, *et al.*, "The unreasonable effectiveness of deep features as a perceptual metric," *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018: 586-595.

[57] K. Simonyan and A. Zisserman, "Very deep convolu-

tional networks for large-scale image recognition," *arXiv preprint* arXiv:1409.1556, 2014.

[58] C. Dong, C. C. Loy, K. He, and X. Tang, "Image super-resolution using deep convolutional networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 2, pp. 295-307, Feb. 2016.

[59] Lim B, Son S, Kim H, *et al.*, "Enhanced deep residual networks for single image super-resolution," *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, 2017: 136-144.

[60] Zhang Y, Li K, Li K, *et al.*, "Image super-resolution using very deep residual channel attention networks," *European Conference on Computer Vision.*, 2018: 286-301.

Sen Lei received his B.S. degree from the Image Processing Center, School of Astronautics, Beihang University in 2015. He is currently working toward his doctorate degree in the Image Processing Center, School of Astronautics, Beihang University. His research interests include deep learning and image super-resolution.

Zhenwei Shi (M¡¯13) received his Ph.D. degree in mathematics from Dalian University of Technology, Dalian, China, in 2005. He was a Postdoctoral Researcher in the Department of Automation, Tsinghua University, Beijing, China, from 2005 to 2007. He was Visiting Scholar in the Department of Electrical Engineering and Computer Science, Northwestern University, U.S.A., from 2013 to 2014. He is currently a professor and the dean of the Image Processing Center, School of Astronautics, Beihang University. His current research interests include remote sensing image processing and analysis, computer vision, pattern recognition, and machine learning.

Dr. Shi serves as an Associate Editor for the *Infrared Physics and Technology*. He has authored or co-authored over 100 scientific papers in refereed journals and proceedings, including the IEEE Transactions on Pattern Analysis and Machine Intelligence, the IEEE Transactions on Neural Networks, the IEEE Transactions on Geoscience and Remote Sensing, the IEEE Geoscience and Remote Sensing Letters and the IEEE Conference on Computer Vision and Pattern Recognition. His personal website is http://levir.buaa.edu.cn/.

Zhengxia Zou received his B.S. degree and his Ph.D. degree from the Image Processing Center, School of Astronautics, Beihang University in 2013 and 2018, respectively. He is now working at the Department of Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, U.S.A., as a postdoc research fellow. His research interests include computer vision, image processing and deep learning. He is a reviewer for the IEEE Transactions on Image Processing, and the IEEE Transactions on Geoscience and Remote Sensing. He has been selected as one of the 2017 best reviewers for the Infrared Physics and Technology. His personal website is http://www-personal.umich.edu/~zzhengxi/.