

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/332089421>

Edge-Enhanced GAN for Remote Sensing Image Superresolution

Article in IEEE Transactions on Geoscience and Remote Sensing · March 2019

DOI: 10.1109/TGRS.2019.2902431

CITATIONS

53

READS

2,082

6 authors, including:



Kui Jiang

Wuhan University

23 PUBLICATIONS 222 CITATIONS

[SEE PROFILE](#)



Zhongyuan Wang

Shanghai University

160 PUBLICATIONS 2,139 CITATIONS

[SEE PROFILE](#)



Peng Yi

Wuhan University

20 PUBLICATIONS 226 CITATIONS

[SEE PROFILE](#)



Guangcheng Wang

Wuhan University

9 PUBLICATIONS 76 CITATIONS

[SEE PROFILE](#)

Some of the authors of this publication are also working on these related projects:



Project Satellite Imagery Superresolution [View project](#)



Project Image Deraining [View project](#)

Edge-Enhanced GAN for Remote Sensing Image Superresolution

Kui Jiang[✉], Zhongyuan Wang[✉], Member, IEEE, Peng Yi[✉], Guangcheng Wang, Tao Lu[✉], and Junjun Jiang[✉], Member, IEEE

Abstract—The current superresolution (SR) methods based on deep learning have shown remarkable comparative advantages but remain unsatisfactory in recovering the high-frequency edge details of the images in noise-contaminated imaging conditions, e.g., remote sensing satellite imaging. In this paper, we propose a generative adversarial network (GAN)-based edge-enhancement network (EEGAN) for robust satellite image SR reconstruction along with the adversarial learning strategy that is insensitive to noise. In particular, EEGAN consists of two main subnetworks: an ultradense subnetwork (UDSN) and an edge-enhancement subnetwork (EESN). In UDSN, a group of 2-D dense blocks is assembled for feature extraction and to obtain an intermediate high-resolution result that looks sharp but is eroded with artifacts and noises as previous GAN-based methods do. Then, EESN is constructed to extract and enhance the image contours by purifying the noise-contaminated components with mask processing. The recovered intermediate image and enhanced edges can be combined to generate the result that enjoys high credibility and clear contents. Extensive experiments on *Kaggle Open Source Data set*, *Jilin-1* video satellite images, and *Digitalglobe* show superior reconstruction performance compared to the state-of-the-art SR approaches.

Index Terms—Adversarial learning, dense connection, edge enhancement, remote sensing imagery, superresolution.

I. INTRODUCTION

IN more recent years, satellite imagery processing has drawn widespread attention because of its special value in extensive application scenarios [1], [2] (e.g., time-span

Manuscript received December 27, 2018; revised February 16, 2019; accepted February 17, 2019. This work was supported in part by the National Natural Science Foundation of China under Grant 61671332, Grant U1736206, Grant 61671336, Grant 61501413, and Grant 61502354, in part the National Key Research and Development Project under Grant 2016YFE0202300, in part by the Hubei Province Technological Innovation Major Project under Grant 2017AAA123, in part by the Central Government Guided Local Science and Technology Development Projects under Grant 2018ZYYD059, and in part by the Basic Research Program of Shenzhen City under Grant JCYJ20170306171431656. (*Corresponding author:* Zhongyuan Wang.)

K. Jiang, Z. Wang, P. Yi, and G. Wang are with the National Engineering Research Center for Multimedia Software, School of Computer Science, Wuhan University, Wuhan 430072, China (e-mail: 2017282110506@whu.edu.cn; wzy_hope@163.com; 2017202110008@whu.edu.cn; wangguangcheng0428@163.com).

T. Lu is with the Hubei Province Key Laboratory of Intelligent Robot, Wuhan Institute of Technology, Wuhan 430205, China (e-mail: lutxyl@gmail.com).

J. Jiang is with the School of Computer Science and Technology, Harbin Institute of Technology, Harbin 150001, China, and also with the Peng Cheng Laboratory, Shenzhen 518055, China (e-mail: jiangjunjun@hit.edu.cn).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TGRS.2019.2902431

comparative studies [3], land cover classification [4]–[7], natural disaster warning [8], assessment of urban economic levels, resource exploration [9], etc.). However, because of hardware cost and craftsmanship limitations, the resolution of the observed images often fails to meet the demand, thus bringing negative impacts on the accuracy of subsequent computer vision tasks [10], [11]. Therefore, how to provide high-quality satellite imageries in a cost-effective manner is mainly discussed in this paper.

Various shallow-learning-based [12], [13] and deep-learning-based reconstruction algorithms have proposed for superresolving satellite imageries in the past years. Especially for the residual learning strategy [14], it was introduced into the computer vision tasks to construct a deeper convolutional neural network (CNN), showing amazing performance. For image superresolution (SR) problems [14]–[16], these methods aim at predicting the residual image (with respect to the input) instead of the target high-resolution (HR) image.

These residual learning-based methods and its variants [17]–[21] have been demonstrated to be effective. However, when the image contents are of over erosion, they will lead to an amount of false or smoothed edges in the resultant images due to the global optimization strategy. Satellite imageries have wide coverage, thus containing a rich variety of ground scenes. The spatial resolution of satellite images is much lower than that of ordinary images. Furthermore, it is subject to additional complex degradation processes, such as ultralong-range imaging, atmospheric disturbances, and equipment noise. Therefore, the conspicuous profile of a ground target is more worth pursuing than the actual texture details within the object. As a result, it further increases the difficulty of recovering the detailed and sharp edges from a given low-resolution (LR) input by the above-mentioned CNNs.

To generate visually releasing results, more recently, generative adversarial network (GAN) [22] has been applied to various computer vision tasks, such as image SR [23], achieving results consistent with the human perception. Adversarial learning strategy [24] is used to push the generative result to the natural image manifold, which is “real” enough for human perception. Although previous works can generate visually pleasing textures, the compensated high-frequency details (e.g., image edges) may be inconsistent with the ground truth, thus resulting in conspicuous artificial artifacts. In particular, the raw GAN-based methods [25] are sensitive to noises,

which tend to generate results including less meaningful high-frequency noise that is irrelevant to the input image. These limitations have negative impacts on the accuracy of high-level computer vision tasks, such as land cover classification [26]–[28] and ground objective recognition.

In order to obtain clear and easily distinguishable detailed edges, researchers further introduce the edge-preserving filters to independently learn image edges. For example, Mao *et al.* [29] proposed a multitask GAN to achieve edge restoration and texture generation simultaneously. They extracted edge structure from the high-quality image as the prior label, which can guide the network to infer the detailed edges. Yang *et al.* [30] improved the SR process with edge-preserving capability based on the fact that LR image and its edge map can jointly infer the sharp edge details of the HR image during the recurrent recovery process. They proposed a deep edge-guided recurrent residual network to progressively realize image SR with properly modeled edge priors, showing sharp high-frequency details. The above-mentioned edge-preserving approaches assume that the prior label of the edge maps is easily inferred. However, when the image contents are of over erosion, they fail to infer and restore the real edges from the noise-contaminated edge priors but lead to noisy results and false image edges. As we presented earlier, the degradation process of satellite imageries is subject to many factors, and the acquired images are often of poor quality compared to the general images. Therefore, previous methods for edge and texture restoration are not suitable for satellite image enhancement.

To address these issues, we propose a GAN-based edge-enhancement network (EEGAN) for video satellite image SR that can generate visually pleasing results while maintaining clean and sharp detailed edges. In EEGAN, the generator consists of two subnetworks: ultradense subnetwork (UDSN) and edge-enhancement subnetwork (EESN). UDSN is used to reconstruct an intermediate SR result that looks sharp but takes noise-contaminated edges. Thus, EESN is specially constructed to enforce the edges by learning a noise mask. More specifically, similar to most dense connection-based methods, UDSN consists of several dense blocks but is used for feature extraction and fusion in a 2-D topology. In the block, each convolutional layer has access to all the subsequent layers both in horizontal and vertical dimensions. In this way, a large number of feature maps are estimated from an LR input through UDSN, followed by a subpixel convolution layer [31] to reconstruct an intermediate SR result. EESN can be roughly divided into five parts: edge extraction, low-dimensional projection, edge enhancement, mask processing, and high-dimensional reconstruction. First, we introduce the Laplacian operator [32] to extract image edges. Then, the convolution layers with integer strides are used to extract feature maps by projecting them onto LR space for releasing the calculation burden. Furthermore, a group of interacted dense blocks in EESN acts to infer and extract the fine edge information. We also construct a mask branch to learn the edge mask and apply it to edge maps to clean up the noise-contaminated image regions. With a subpixel upsampling operation, the enhanced edge maps are then transformed into the HR space. At last,

we replace the sharp but dirty edge maps in the intermediate SR result with the enforced ones and reconstruct an SR image with sharp edges and noise-free contents.

In summary, the major contributions of this paper are at least threefold.

- 1) We propose an EEGAN for video satellite image SR reconstruction in a convenient and effective end-to-end training manner. EEGAN performs favorably against state-of-the-art CNN- and GAN-based methods.
- 2) We propose a novel approach to simultaneously infer and estimate the perceptually pleasing textures and easily discernible contours from an LR image through constructing different subnetworks and optimizing each branch with special objectives.
- 3) An effective and practical edge-enhancement method is proposed to enhance image contours, which can effectively remove the noises while maintaining sharp and detailed edges. In particular, we put forward a mask branch to reduce the artificial artifacts and noises produced by the adversarial learning through purifying the noise-contaminated image regions.

The remainder of this paper is organized as follows. In Section II, we introduce previous works related to our study, such as image SR reconstruction, adversarial learning, edge extraction, and dense connection. Section III particularly presents the proposed EEGAN framework. We also individually present the design details of each key module under the EEGAN framework. Experimental results are given in Section IV, and Section V concludes this paper.

II. RELATED WORK

Deep CNNs [33]–[36] have been widely applied to the image SR reconstruction. A comprehensive review is beyond the scope of this paper and we discuss the most related ones in this section.

A. Image Superresolution

In recent years, a variety of SR techniques for the restoration of HR remote sensing images have been proposed [37]–[39]. For example, Merino and Nunez [13] proposed a variable-pixel linear reconstruction algorithm, which recombines a set of LR images in a linear nonuniform optimum manner. In [12], a hidden Markov tree model was proposed to establish a prior model in the wavelet domain to regularize the ill-conditioned problem for remote sensing image SR restoration. However, the shallow-learning-based models show poor performance on nonlinear degradation, whereas satellite imaging is a typical nonlinear process due to diversified complex factors. Recently, CNNs have shown powerful capability in promoting reconstruction quality. SRCNN [40], as a pioneering deep-learning based framework, learns a mapping relationship from LR to HR with a three-layer CNN structure by an end-to-end manner. Again, Dong *et al.* [41] proposed Fast Super-Resolution Convolutional Neural Networks (FSRCNN) to reduce the reconstruction time by replacing the bicubic interpretation with the LR input and strengthen the feature extraction with more hidden layers. Lei *et al.* [42] constructed

a multifork framework to learn a multilevel representation of remote sensing images, including both local details and global environmental priors.

B. Generative Adversarial Network

GAN [29], [43] is another popular deep learning model after CNN, which consists of a generative network and a discriminative network. Generally, the generative network is trained to generate samples that keep in line with real data in manifold, whereas the discriminative network is trained to determine whether a sample is from real data or the generator. During the training, the two networks compete with each other so that the distribution captured from the generator is as similar as possible to the distribution of the real data.

Since GAN was proposed by Goodfellow *et al.* [22], researchers have studied it vigorously. Recently, several effective and practical techniques have been applied to low-level computer vision tasks, including image SR [44], [45]. For example, Ledig *et al.* [25] proposed a photorealistic single-image SR using a GAN (SRGAN), which uses adversarial loss to push the reconstruction result to the natural image manifold. Sajjadi *et al.* [23] conducted an EnhanceNet to synthesize realistic textures based on adversarial training, perceptual losses, and texture transfer loss. Recently, Wu *et al.* [46] proposed a robust perceptual loss to encourage the generator to reconstruct photorealistic SR results. Usually, these methods generate visually pleasing textures by measuring the perceptual similarity in the feature level.

C. Edge Extraction Operator

To the best of our knowledge, only a few approaches have been proposed to address image edges extraction and enhancement problems. Xu *et al.* [47] developed a CNN model to approximate a number of filters by learning a family of edge-aware operators from data. Liu *et al.* [48] used a hybrid CNN to approximate edge-preserving filters to extract and preserve the image edges. These methods aim to preserve the important structures and remove details using a single network but impose a difficult learning task. Alonso *et al.* [49] presented a robust and unsupervised edge-enhancement algorithm by combining wavelet coefficients at different scales. Bao *et al.* [50] defined the detection and localization criteria of the scale multiplication. Through scale multiplication, the localization accuracy can be significantly improved, thus leading to superior edge detection results. However, the existing methods fail to accurately estimate both the edges and details of the reconstructed image.

D. Dense Connection

The skip connection was first introduced for image SR by Kim *et al.* [15], who constructed a very deep CNN for accurate image SR (VDSR). Instead of learning the actual pixel values, the authors used the residual learning paradigm to predict the differences between the HR and the bicubic interpolated image, which makes the feature maps very sparse, enabling easy training and convergence. More recently, a densely connected strategy [18], [51], [52] is also adopted for image reconstruction by connecting the feature maps of the current

layer to every subsequent layer in a feed-forward manner. For example, Tai *et al.* [53] integrated recursive learning and skip connections for image restoration tasks. They proposed long-term dense connections to recover much more high-frequency information. Huang *et al.* [52] exploited densely connected convolutional layers to maintain coarse and fine feature classification through a multi-scale network.

III. METHOD

In this paper, our ultimate goal is to learn a mapping function F that can reconstruct the corresponding HR image from the given LR input. As illustrated in Fig. 1, our proposed framework EEGAN is composed of three main components: a generator (G), a discriminator (D), and a VGG16 network for feature extraction. The generator G can be roughly partitioned into two substructures: an UDSN and an EESN. UDSN is composed of several dense blocks and a reconstruction layer for generating an intermediate HR result. EESN is used to enhance the target edges extracted from the intermediate SR image by cleaning up the noises and artifacts. We obtain the final SR output by replacing the noisy edges in the intermediate SR image with the purified edges from EESN. During training, the generator G tries to transform the input LR image into HR image space similar to the ground truth, returning two reconstruction results: an intermediate SR result with sharp but noise-contaminated and dirty contents and the final SR result with sharp contours and clean contents. For the discriminator D , we take the architectural design in [25] as a reference but use the maximum pooling to replace the strided convolution. It encourages the generator G to generate a realistic image more faithful to the ground truth. By training and updating these two networks iteratively, it converges to an HR output faithful to the ground truth.

A. Model Optimization

In this paper, I_{LR} , I_{Base} , I_{SR} , and I_{HR} are considered as the input, the intermediate SR image, the SR output, and the ground truth of the proposed EEGAN, respectively. We let I_{Edge} and I_{Edge}^* denote the intermediate and enhanced edge maps from I_{Base} and EESN, respectively. We construct a robust content loss function to enforce the generator G to generate a intermediate HR image I_{Base} similar to I_{HR} using the following model:

$$\text{Loss}_{\text{cont}}(\theta_G) = \arg \min_{\theta_G} \sum_{i=1}^n \rho(I_{HR,i} - I_{Base,i}) \quad (1)$$

where θ_G denotes a set of model parameters in G , and $\rho(x) = (x^2 + \varepsilon^2)^{(1/2)}$ represents the Charbonnier penalty function [33]. We empirically set the compensation parameter ε of 10^{-3} . $I_{Base,i}$ and $I_{HR,i}$ refer to the extracted features of the reconstructed image and the ground truth by VGG-net, respectively.

Then we feed I_{Base} and I_{HR} into D to determine their authenticities, real or fake. We train the discriminator by minimizing the adversarial loss, which can encourage G to generate the reconstructed image I_{SR} close to the ground truth I_{HR} . This process can be formatted as follows:

$$\text{Loss}_{\text{adv}}(\theta_D) = -\log D(I_{HR}) - \log(1 - D(G(I_{LR}))) \quad (2)$$

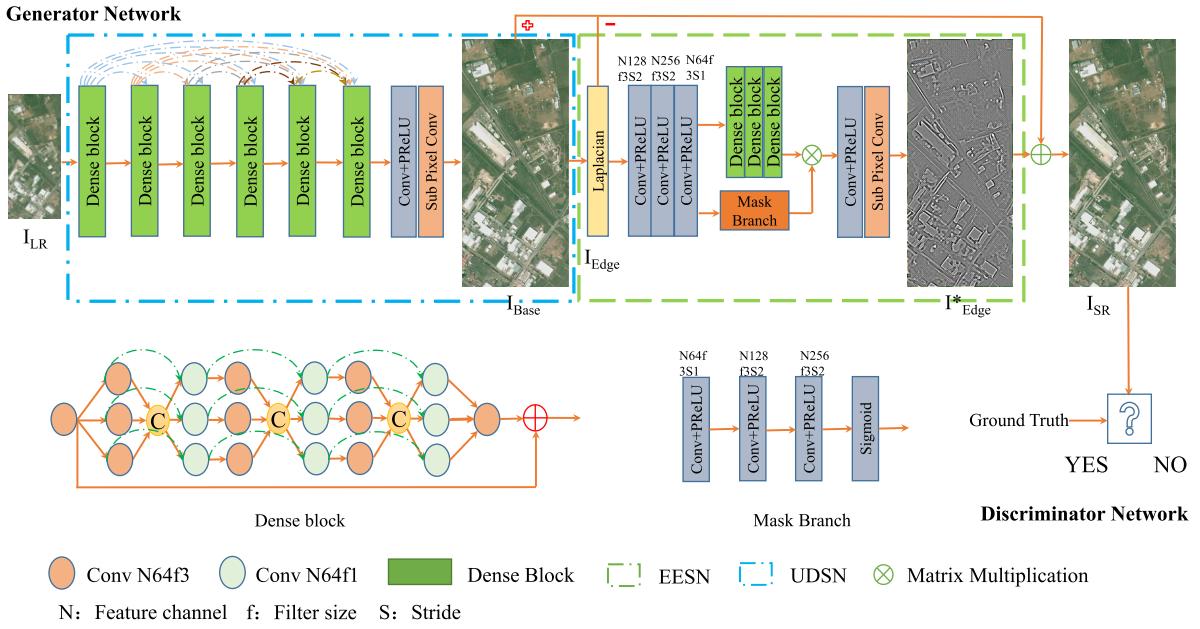


Fig. 1. Outline of the proposed edge-enhancement network (EEGAN). I_{LR} , I_{Base} , and I_{SR} denote the input, the intermediate SR image, and the final SR output of the proposed EEGAN, respectively. I_{Edge} and I_{Edge}^* represent the intermediate edge maps extracted by the Laplacian operator and the enhanced edge maps from EESN.

where θ_D refers to the model parameters in D_{net} . I_{LR} denotes the LR input. $G(\cdot)$ represents the function of the generative network, while $D(\cdot)$ is the discriminative function to calculate the probability whether the input is the real HR image or fake superresolved image.

It is well known that adversarial learning strategies can maintain the visual authenticity of the generated images but with a large number of artifacts and indeterminate details. In order to improve the quality of reconstructed image and reduce artifacts, we further introduce the pixel-based Charbonnier loss to enhance the consistency of image contents between I_{SR} and I_{HR} . The consistency loss function is formulated as follows:

$$\text{Loss}_{\text{cst}}(\theta_G) = \rho(I_{HR} - I_{SR}). \quad (3)$$

where θ_G denotes the model parameters in G . I_{SR} and I_{HR} refer to the final SR output and the ground truth, respectively. Finally, the total objective is given by

$$\begin{aligned} L(\theta_G, \theta_D) = & \text{Loss}_{\text{cont}}(\theta_G) + \alpha \text{Loss}_{\text{adv}}(\theta_G, \theta_D) \\ & + \lambda \text{Loss}_{\text{cst}}(\theta_G) \end{aligned} \quad (4)$$

where α and λ are the weight parameters to balance the loss components. We empirically set the weight parameters α and λ to 1×10^{-3} and 5.

B. Feature Extraction and Stage Reconstruction

CNNs have been widely studied for feature extraction in recent years, such as the highway network [54], skip connection [55], dense-connection network [51], and memory network [56]. These typical and effective frameworks have gained amazing performance in various computer vision tasks. Especially for the recently popular dense-connection manner,

it promotes the information sharing and fusion by linking the current layer to all subsequent layers with numerous skip connections. Various practical applications [18], [51] have verified its superiority in feature extraction and expression. In this paper, we have constructed a novel dense residual block with 2-D topology, which is served as one of the basic components for feature extraction and expression. Benefiting from the ingenious designing and effective structures, the feature maps extracted from multiple paths can be shared better through rich diagonal connections. When compared to the conventional dense residual block, the additional diagonal connections between different paths in this 2-D structure can provide more possibilities for the transformation of information flow and better gradient optimization. To put it in another way, the features can be additionally established through the diagonal connections so that more and enough link pathways can be formed with the same number of layers. Furthermore, the sharing mechanism of 2-D local features in the ultradense block increases the flow of information, thus resulting in reasonable feature extraction and expression during training.

As shown in Fig. 1, we regard the ultradense block as the basic module for feature extraction and fusion both in UDSN and EESN. Different from the traditional dense block [18], [51], the subsequent layer in our work can share and fuse the feature maps extracted from multiple preceding convolution layers both in the horizontal and vertical dimensions. Therefore, there are about twice as many linking nodes as original dense block does, leading to diversified and fine feature expression. This procedure in our framework can be formulated as

$$y_i = H_C([F_{i,0}(x_0), F_{i,1}(x_1), \dots, F_{i,n}(x_n)]) \quad (5)$$

$$s_{i,n} = H_1(H_C(y_i, s_{i-1,n})). \quad (6)$$

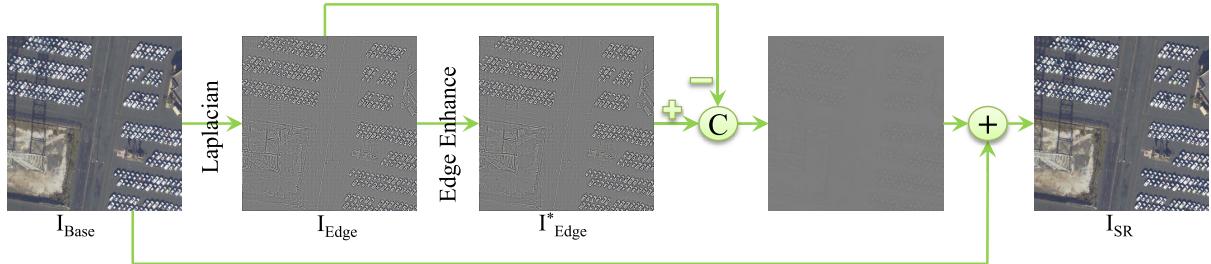


Fig. 2. Outline of the proposed edge enhancement. I_{Edge} and I_{Edge}^* refer to the edge maps extracted from the intermediate and final SR results, respectively.

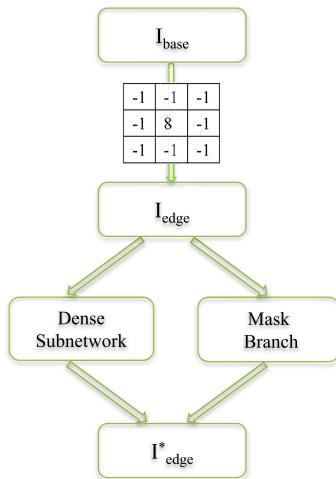


Fig. 3. Outline of the proposed subnetwork EESN.

In (5), x_n refers to the input of the n th path in the 2-D dense block. $F_{i,n}(x_n)$ denotes a single convolution operation of the n th path in the i th dense block with the filter size of 3×3 . $H_C([F_{i,0}(x_0), F_{i,1}(x_1), \dots, F_{i,n}(x_n)])$ refers to the feature congregation of multiple layers in the vertical dimension in the dense unit, and y_i denotes the fused features. In (6), $s_{i,n}$ indicates the transition output for obtaining the equivalent feature channels through a 1×1 filter. Then we feed the features into subsequent blocks for further extraction and fusion. Our effective sharing strategies between ultradense blocks greatly benefit to the training.

Finally, following the last block, a subpixel reconstruction layer [31] is used to obtain the superresolved image by projecting the feature maps in LR space onto HR space. To achieve the amplification operation without traditional interpolation, we learn $C \times r^2$ feature maps for reconstructing an SR image with the subpixel convolution. It can convert the feature maps in the size of $W \times H \times Cr^2$ into an RGB image in the size of $Wr \times Hr \times C$. Thus, an intermediate HR result $I_{\text{Base}} \in R^{rH \times rW} \times C$ is obtained from the given LR image $I_{\text{LR}} \in R^{H \times W} \times C$. This operation can be described as follows:

$$I_{\text{Base}} = \text{PS}(F(I_{\text{LR}})) = \text{PS}(W_n \otimes F_{n-1}(I_{\text{LR}}) + b_n) \quad (7)$$

where $F_{n-1}(I_{\text{LR}})$ denotes the output of the $(n-1)$ th layer, W_n and b_n represent the weight and bias of the last layer, and \otimes refers to the matrix multiplication. $F(\cdot)$ refers to the convolution operation, followed by the leaky rectified linear unit (Leaky-ReLU) for activation.

C. Edge Extraction and Enhancement

It is generally known that the generative adversarial model can push the generator to generate results keeping in line with the true manifold but also add less meaningful high-frequency noise that is irrelevant to the input image. To address this issue, we design an effective subnetwork EESN to extract edge features that are useful for identification. The EESN consists of a dense subnetwork and mask branch, where the dense subnetwork is utilized to extract fine edge maps. However, the edge maps extracted from the base image are noisy and will cause difficulties for subsequent discrimination. Therefore, we simultaneously construct a mask branch to learn the noise mask through the attention mechanism so that the network can focus on the real edge information to achieve the purpose of removing noises and artifacts. More specifically, a mask branch adaptively learns specific weight matrices with soft attention to the relevant information.

As shown in Fig. 2, the edge extraction and enhancement module takes the stage results I_{Base} obtained from UDSN as input. EESN needs to label each pixel in a binary variable, indicating whether the pixel contains an edge or not. As a typical approach, Laplacian operator can highlight regions of rapid intensity change and is, therefore, widely adopted in previous works for edges detection and extraction [32], [57], [58]. Laplacian is, thus, utilized in our framework to label the image edges over the SR images I_{Base} .

For the given samples from UDSN, we use Laplacian operator [32] to detect and extract the edges. The Laplacian $L(x, y)$ of an image $I(x, y)$ can be defined as its second derivatives, which is formulated as follows:

$$L(x, y) = \frac{\partial^2 I}{\partial x^2} + \frac{\partial^2 I}{\partial y^2}. \quad (8)$$

Laplacian operator possesses isotropy and rotation invariance, which makes the image produce a steep zero-crossing point at the edge after the second derivatives. According to the zero-crossing point given by (8), the edge can be determined. In particular, we use $([-1, -1, -1], [-1, 8, -1], [-1, -1, -1])$ as the discrete convolution mask of the Laplacian. Using the convolution kernel, Laplacian procedure is given as

$$E(x, y) = L(x, y) \otimes I(x, y) \quad (9)$$

where $L(x, y)$ refers to the discrete convolution mask and $E(x, y)$ represents the extracted edge map. \otimes is the convolution operator.

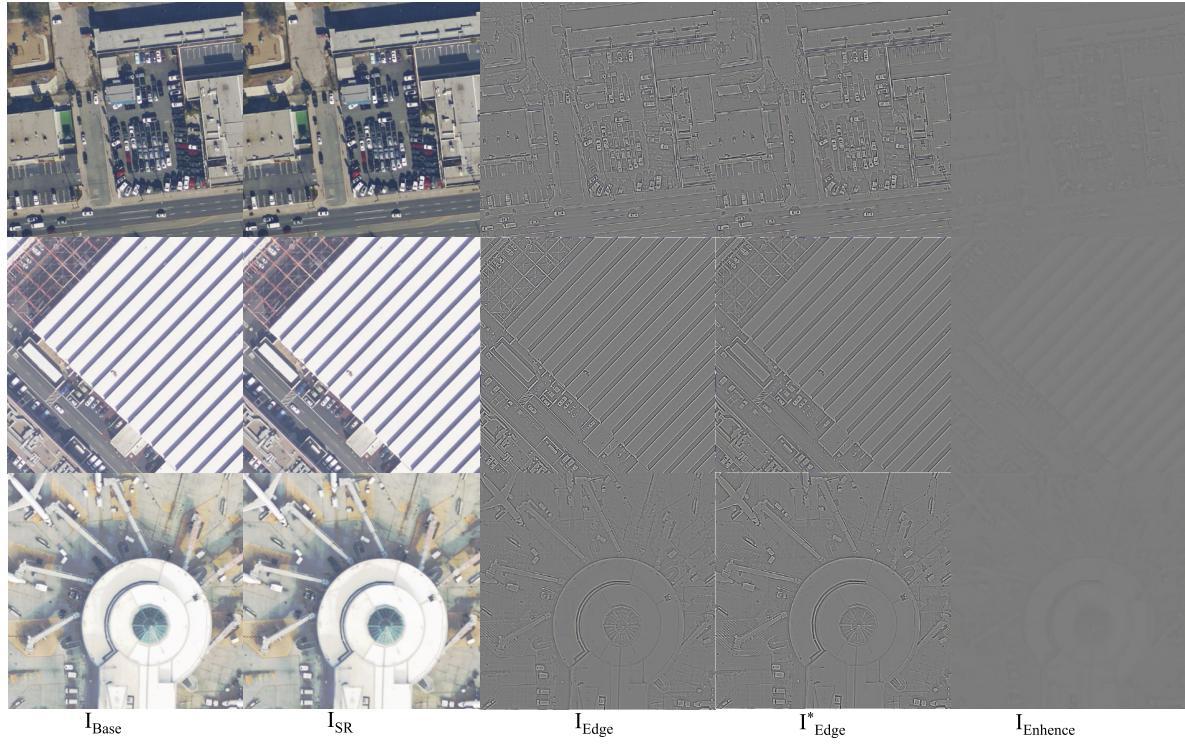


Fig. 4. Image reconstruction results of different stages in the proposed method. I_{Edge} and I^*_{Edge} refer to the edge maps extracted from the intermediate and final SR results, respectively. I_{Enhance} denotes the added details.

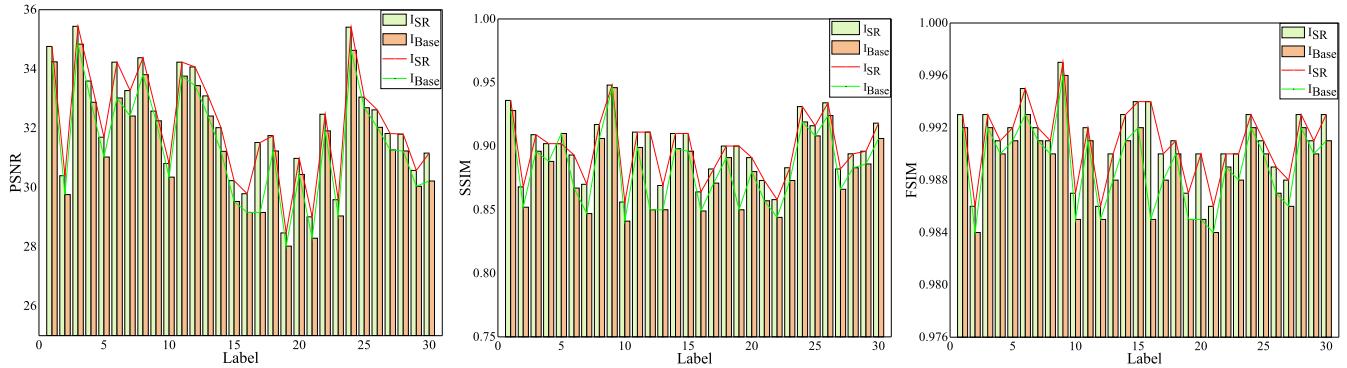


Fig. 5. Comparison results of PSNR, SSIM, and FSIM on Kaggle Open Source data set between I_{Base} and I_{SR} by the scale factor of 4.

Inspired by Ledig *et al.* [25], we utilize the strided convolution to extract edge maps and simultaneously transform them into LR space. In contrast to the operation in HR space, this strategy can reduce the amount of calculation. In addition, we also utilize a similar topological structure in UDSN to estimate and enhance the edge information in EESN. At the same time, as shown in Fig. 3, a mask branch is conducted to learn the image mask to detect and clean up the isolated noises that are the false edge points produced in edge extraction. Subsequently, the enhanced edge maps are projected onto HR space by a subpixel convolution operation. The above-mentioned operations in EESN can be formulated as follows:

$$I^*_{\text{Edge}} = \text{PS}(F(D(I_{\text{Edge}})) \otimes M(D(I_{\text{Edge}}))) \quad (10)$$

where $D(\cdot)$ represents the downsampling operation with the strided convolution. It projects the edge maps I_{Edge} extracted by the Laplacian operator onto LR space. In (10), $F(\cdot)$ refers

TABLE I
COMPARISON RESULTS OF ABLATION STUDY ON KAGGLE OPEN SOURCE DATA SET WITH THE SCALE FACTOR OF 4.

Model	EENAN_BASE	EENAN_NMASK	EENAN_NUD	EENAN
PSNR	31.75	31.85	32.00	32.36
SSIM	0.886	0.889	0.893	0.898
FSIM	0.989	0.990	0.990	0.991

to the dense subnetwork in EESN using for feature extraction and fusion. $M(\cdot)$ denotes the mask branch to learn the image mask to suppress the noises and the false edges. $PS(\cdot)$ is the upsampling operation with subpixel convolution to transform the edge maps into HR space. At the end of the network, we replace the noise-contaminated edges in I_{Base} with the enhanced edge maps I^*_{Edge} to obtain the HR reconstruction result. We illustrate the procedure in Figs. 2 and 3.

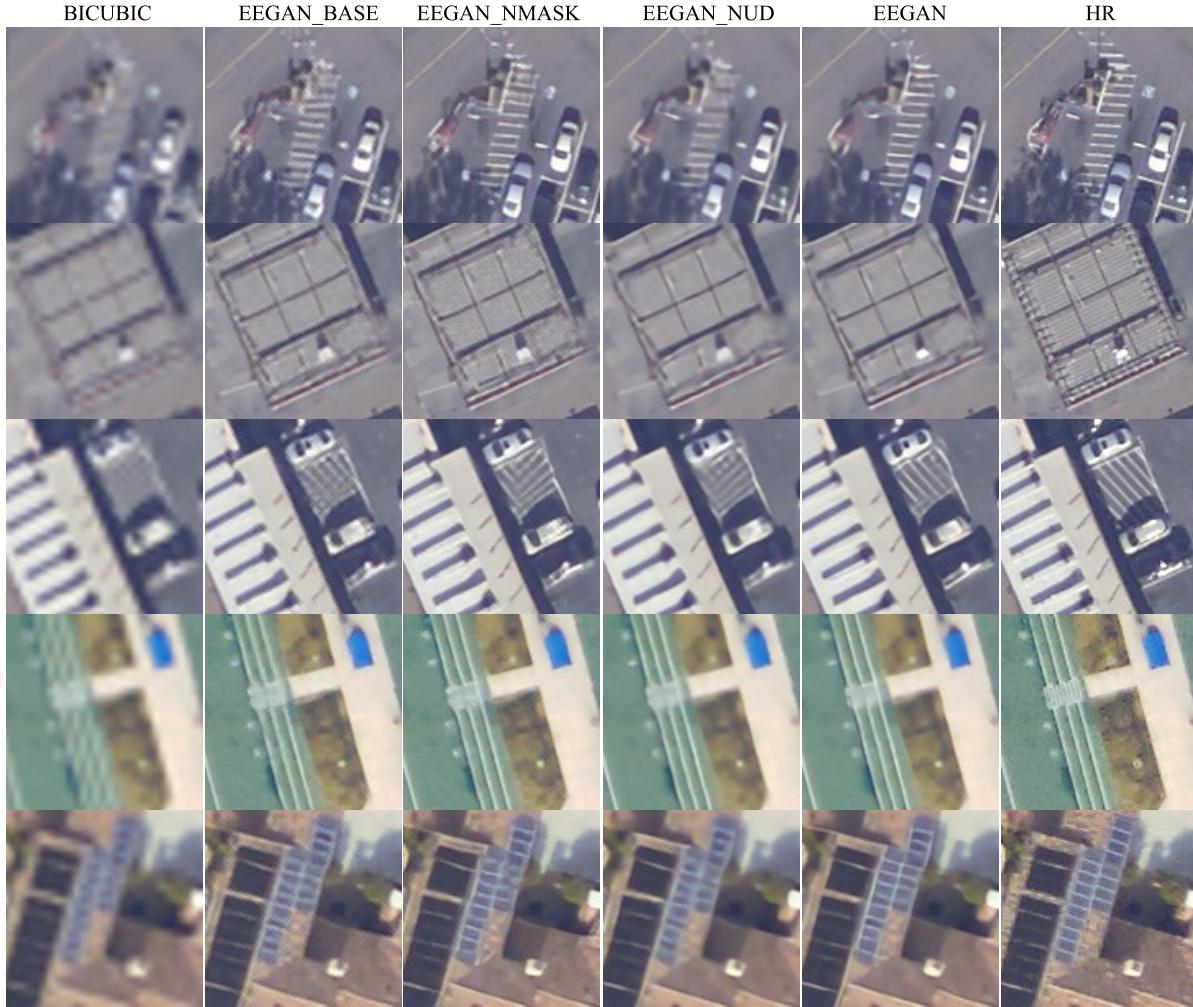


Fig. 6. Reconstruction results on Kaggle Open Source data set by the scale factor of 4.

TABLE II

COMPARISON RESULTS OF AVERAGE PSNR, SSIM, AND FSIM ON KAGGLE OPEN SOURCE DATA SET WITH SCALE FACTORS OF 2, 3, AND 4.
* DENOTES THAT THE MODELS ARE RETRAINED WITH OUR TRAINING DATA SET

Methods	Bicubic	SRCNN* [41]	VDSR* [15]	SRGAN* [25]	EEGAN (Ours)
Metrics	PSNR/SSIM/FSIM	PSNR/SSIM/FSIM	PSNR/SSIM/FSIM	PSNR/SSIM/FSIM	PSNR/SSIM/FSIM
$\times 2$	34.01/0.938/0.997	36.79/0.960/0.998	37.94/0.967/0.999	37.69/0.963/0.998	38.82/0.973/0.999
$\times 3$	30.52/0.870/0.985	32.44/0.906/0.994	33.69/0.924/0.996	33.70/0.919/0.993	34.84/0.936/0.995
$\times 4$	28.54/0.808/0.966	30.06/0.848/0.985	31.06/0.874/0.989	31.17/0.882/0.986	32.36/0.898/0.991

IV. EXPERIMENTS AND DISCUSSION

In this section, we compare our method with representative SR methods, including bicubic, SRCNN [40], VDSR [15], and SRGAN [25]. Similar to many previous works [53], [59], we also adopt several commonly used evaluation metrics (i.e., peak-signal-to-noise ratio (PSNR), feature similarity (FSIM) [60], and structural similarity (SSIM) [61]) for comparison. These evaluation metrics differ in visual perception and, thus, give a comprehensive comparison, but reference images are needed. In addition, we introduce the indicators without reference, including average gradient (AG) [62] and naturalness image quality evaluator (NIQE) [63], to evaluate the real-world SR reconstruction without HR reference.

A. Data Set

In this paper, two available satellite image data sets, namely, *Kaggle Open Source Data set*¹ and *Jilin-1* video satellite imageries are used in our experiments. For Kaggle Open Source Data set, it consists of subjects with large diversities, large quantities, and rich annotations, including 324 different scenarios with a size of 3099×2329 pixels. Its 1720 (350 for training and 1370 for testing) satellite images cover agriculture, airplane, buildings, golf course, forest, freeway, parking lot, tennis court, storage tanks, harbor, and so on. The training samples are cropped into small image batches (96×96) for training without overlapping for any scaling factors. For the

¹<https://www.kaggle.com/c/draper-satellite-image-chronology/data>

TABLE III
COMPARISON RESULTS OF PSNR, SSIM, AND FSIM ON KAGGLE OPEN SOURCE DATA SET BY THE SCALE FACTOR OF 4.

Labels	Methods Scale	Bicubic PSNR/SSIM/FSIM	SRCNN* [41] PSNR/SSIM/FSIM	VDSR* [15] PSNR/SSIM/FSIM	SRGAN* [25] PSNR/SSIM/FSIM	EEGAN (Ours) PSNR/SSIM/FSIM
(1)	4	30.84/0.868/0.975	32.42/0.898/0.988	33.60/0.920/0.992	33.45/0.924/0.990	35.12/0.939/0.994
(2)	4	26.41/0.740/0.950	27.77/0.788/0.976	29.07/0.833/0.985	29.48/0.851/0.980	30.49/0.867/0.985
(3)	4	32.36/0.852/0.972	33.82/0.879/0.989	34.45/0.896/0.992	34.49/0.898/0.990	35.44/0.908/0.993
(4)	4	30.52/0.848/0.971	32.13/0.876/0.987	33.02/0.893/0.991	32.26/0.888/0.987	33.78/0.903/0.992
(5)	4	25.31/0.816/0.962	27.75/0.867/0.986	30.73/0.907/0.992	30.63/0.909/0.989	32.17/0.922/0.993
(6)	4	29.58/0.800/0.973	30.52/0.825/0.989	30.73/0.907/0.992	32.99/0.876/0.991	34.26/0.891/0.995
(7)	4	31.04/0.822/0.979	32.20/0.845/0.990	32.80/0.859/0.992	32.99/0.876/0.991	33.30/0.868/0.993
(8)	4	31.58/0.872/0.976	32.85/0.894/0.989	33.61/0.907/0.992	32.99/0.876/0.991	34.43/0.918/0.991
(9)	4	26.91/0.831/0.958	29.95/0.898/0.989	31.64/0.935/0.995	32.01/0.944/0.994	32.56/0.950/0.997
(10)	4	28.47/0.783/0.967	29.38/0.817/0.983	29.94/0.835/0.987	30.16/0.842/0.982	30.84/0.858/0.987
(11)	4	31.32/0.858/0.973	33.04/0.888/0.990	33.87/0.903/0.993	33.42/0.901/0.989	34.19/0.910/0.993
(12)	4	32.50/0.831/0.974	33.53/0.854/0.987	33.97/0.863/0.989	33.07/0.855/0.983	33.91/0.865/0.986
(13)	4	30.75/0.823/0.976	32.04/0.854/0.989	32.68/0.865/0.991	31.88/0.855/0.986	32.90/0.865/0.990
(14)	4	27.94/0.830/0.973	29.60/0.865/0.987	30.55/0.886/0.991	30.80/0.896/0.988	32.40/0.915/0.994
(15)	4	25.71/0.744/0.965	27.29/0.808/0.988	28.28/0.845/0.993	29.33/0.891/0.989	30.22/0.914/0.994
(16)	4	25.98/0.738/0.960	27.29/0.796/0.981	28.16/0.830/0.986	28.72/0.846/0.982	30.19/0.870/0.988
(17)	4	27.91/0.785/0.966	29.51/0.830/0.985	30.59/0.861/0.990	30.72/0.871/0.986	31.83/0.886/0.991
(18)	4	28.11/0.811/0.969	29.73/0.853/0.986	30.86/0.883/0.990	30.94/0.891/0.988	32.08/0.904/0.992
(19)	4	25.80/0.734/0.961	27.00/0.796/0.982	27.43/0.821/0.986	27.81/0.845/0.983	28.82/0.872/0.988
(20)	4	27.07/0.767/0.962	28.69/0.822/0.983	29.82/0.862/0.989	30.04/0.877/0.986	31.36/0.897/0.991
(21)	4	24.87/0.734/0.948	26.32/0.794/0.975	27.50/0.837/0.984	28.04/0.856/0.981	29.45/0.881/0.988
(22)	4	30.73/0.812/0.976	31.92/0.840/0.989	32.38/0.853/0.991	31.65/0.847/0.987	32.42/0.857/0.990
(23)	4	25.85/0.779/0.961	27.73/0.833/0.984	28.60/0.861/0.989	28.83/0.872/0.986	29.98/0.891/0.991
(24)	4	32.16/0.884/0.974	33.92/0.906/0.989	34.73/0.921/0.992	34.16/0.918/0.990	35.37/0.932/0.994
(25)	4	29.08/0.839/0.963	30.38/0.872/0.983	31.80/0.899/0.989	32.29/0.906/0.988	33.01/0.917/0.991
(26)	4	27.96/0.824/0.945	29.02/0.862/0.974	29.59/0.883/0.981	30.90/0.912/0.980	33.88/0.943/0.990
(27)	4	29.06/0.824/0.971	30.26/0.855/0.985	31.41/0.874/0.989	31.06/0.871/0.984	31.69/0.880/0.988
(28)	4	28.34/0.801/0.974	29.88/0.844/0.989	30.65/0.869/0.99	30.94/0.884/0.990	32.22/0.900/0.994
(29)	4	26.81/0.783/0.963	28.56/0.837/0.986	29.53/0.870/0.990	29.84/0.885/0.987	31.01/0.903/0.992
(30)	4	25.46/0.810/0.960	27.35/0.862/0.985	29.60/0.902/0.990	29.90/0.905/0.989	31.75/0.923/0.993
Avg	4	28.54/0.808/0.966	30.06/0.848/0.985	31.06/0.874/0.989	31.17/0.882/0.986	32.36/0.898/0.991

test images, the central part with a size of 720×720 is cropped as the candidate. For Jilin-1 video satellite imageries, we select one for every five frames from each video and crop the central part with the size of 480×204 as test samples. Several areas in different countries with certain typical surface coverage types, including vegetation, harbor, and a variety of buildings, are tested.

B. Experimental Details

In our two subnetworks, the depths of ultradense block are set to 6 and 3, respectively. The filter size and depth of each layer are set to 3×3 and 64, followed by the Leaky-ReLU activated function. For training, the input patches are cropped in a size of 96×96 pixels with a batch size of 16. The learning rate is initialized as 2×10^{-4} with a reduced rate of 0.9 after every epoch till 1×10^{-5} .

We use an NVIDIA Titan XP GPU and an Intel I7-8700 CPU for training and testing, respectively. Our model is implemented on TensorFlow with Python3 under Windows10, CUDA8.0, and CUDNN5.1 systems. In experiments, training an EEGAN roughly takes 20 h with a Titan XP GPU for 25 epochs on the above-mentioned settings.

C. Validation of the Edge Enhancement

To demonstrate the effectiveness of the proposed edge-enhancement strategy along with the adversarial learning, in this section, we show the comparison results. By performing the ultradense image SR reconstruction, we obtain

an intermediate HR reconstruction result I_{Base} , which can well maintain the primary image contents but with noise-contaminated edges and artifacts. Through Laplacian operator, we detect and extract the edge maps, though containing false edge points (noises and artifacts). Then, a dense subnetwork and a mask branch are embedded to gradually enhance the details of the edge maps and obtain the enhanced edge maps I_{Edge}^* . As shown in Fig. 4, by applying the image mask to the edge maps, the noises and artifacts are well mitigated. Fig. 5 quantitatively shows the effectiveness of the proposed edge-enhancement strategy with a significant improvement in all indicators (PSNR, SSIM, and FSIM). This validation confirms that edge enhancement is helpful to produce realistic SR results.

We further conduct ablation study on the effects of the proposed EESN. Since our edge-enhancement strategy has several incorporative components, we clearly show how the performance improves with different parts based on the performance of the baseline. In this test, we conduct three experiments to estimate the ultradense module and the mask branch, respectively. We refer to our final model as EEGAN, which is composed of UDSN and EESN. In each experiment, we further compare our final model EEGAN with three other network structures. Specifically, by removing the EESN from EEGAN, the remaining parts constitute the first network, named EEGAN_BASE. The second network, named EEGAN_NMASK, can be regarded as an enhanced variant of EEGAN_BASE with an added UDSN for edge feature extraction. When compared with

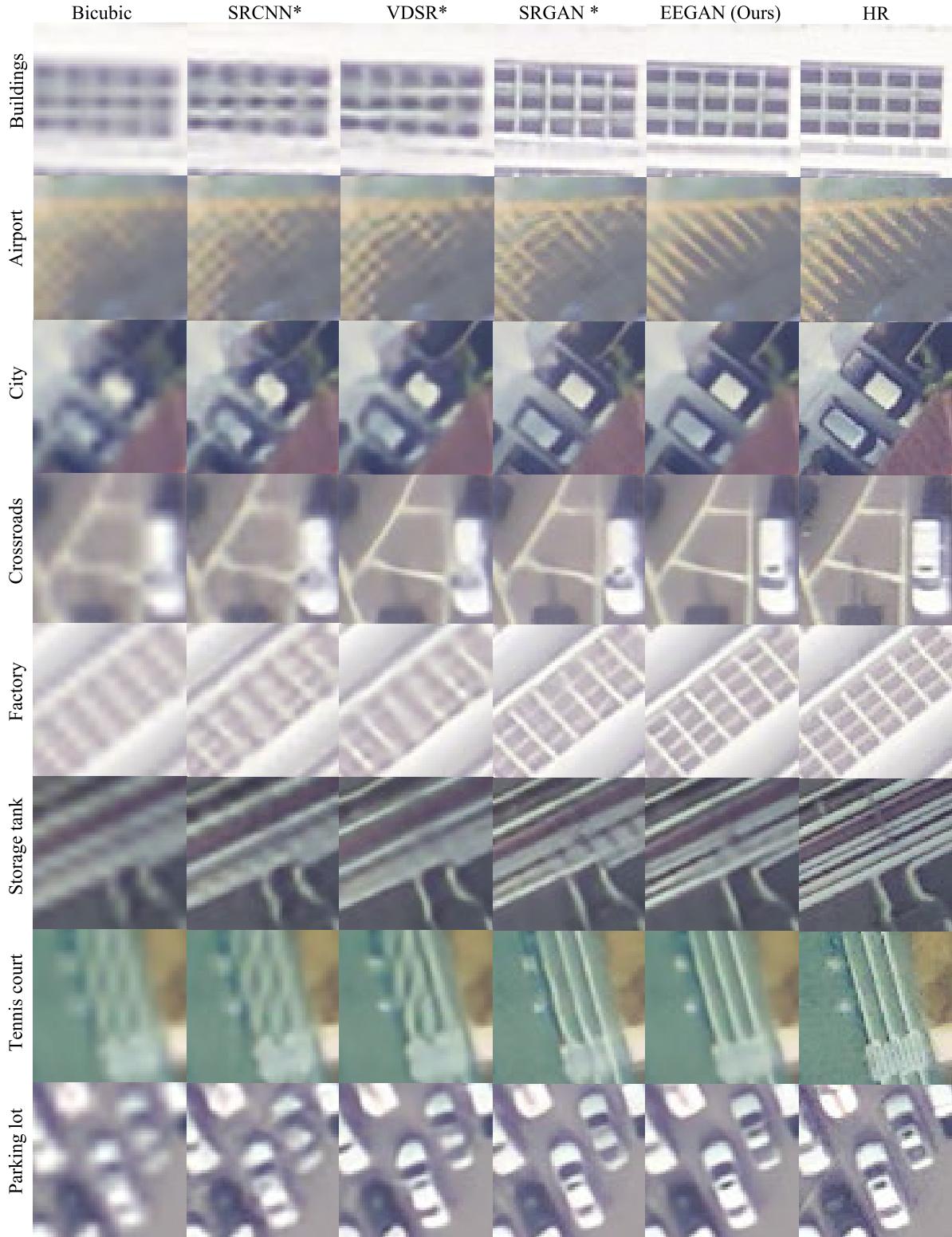


Fig. 7. Reconstruction results on Kaggle Open Source data set by the scale factor of 4. We select several different but representative scenarios, i.e., crossroads, factory, freeway, tennis court, and parking lot.

EEGAN_BASE, EEGAN_NUD additionally learns a mask to further clean up noises and artifacts through a CNN subnetwork.

The comparison results in terms of PSNR, SSIM, and FSIM are reported in Table I. In addition, we also compare the

visual results of different models, as shown in Fig. 6. From these scores, it is obvious that the basic bicubic interpolation method cannot produce enriched details. Our basic model (EEGAN_BASE) can generate sharp image edges but with dirty contents. Simply increasing the depth of the model



Fig. 8. Reconstruction results on Jilin-1 data set with the scale factor of 4. Several different but representative scenarios, i.e., factory, airport, and city suburbs, are used for comparison.

for edge feature extraction (EEGAN_NMASK) can slightly improve the reconstruction performance by inferring more fine details while still having the amount of noises and artifacts. By introducing the noise mask strategy (EEGAN_NUD), the noises and artifacts generated by the adversarial training can be well removed in addition to less high-frequency information. In particular, by incorporating the ultradense connection and mask mechanism, the reconstruction results show a significant improvement with high credibility and clear contents, approximately surpassing the conducted EEGAN_BASE 0.61 dB with the scale factor of 4.

D. Comparisons With State of the Arts on Kaggle Open Source Data Set

To illustrate the practicability, we compare our method with several representative CNN- and GAN-based SR methods, including SRCNN [40], VDSR [15], and SRGAN [25].

For a fair comparison, we retrain these models using the same training data sets and label them as SRCNN*, VDSR*, and SRGAN*.

As shown in Table II, EEGAN exhibits the highest scores in all indicators, including PSNR, SSIM, and FSIM, compared with CNN-based (SRCNN and VDSR) or GAN-based (SRGAN) methods. On average, the proposed EEGAN surpasses SRCNN, VDSR, and SRGAN about 2.30, 1.30, and 1.19 dB when the amplification factor is set to 4. The evaluation results of the scale factor of $\times 4$ are additionally reported in Table III in detail.

As shown in Fig. 7, we also compare the visual results of different methods. It shows that the basic bicubic interpolation method cannot produce extra details. As for the deep learning-based technologies, such as SRCNN and VDSR, they can infer some texture details but result in blurry image contours due to their global optimization scheme and poor utilization of features (refer to the outlines of the buildings, roads,



Fig. 9. Reconstruction results on Digitalglobe with the scale factor of 4. We select the images of Solar power plant and Burj Dubai for comparison.

cars, etc.). When compared with SRGAN, which can be seen as the current most competitive GAN-based method for image SR, our results are still very competitive and much more realistic. SRGAN exhibits a number of artificial artifacts that do not exist in ground truth, leading to dirty reconstruction results.

E. Comparisons With State of the Arts on Jilin-1 Video Satellite

Jilin-1 video satellite is in 1.12-m resolution, which is far below that of traditional still remote sensing imageries. Compared with Kaggle Open Source data set, the samples obtained from Jilin-1 take a lower quality, for example, smaller ground objects and weaker textures. For this reason, SR for satellite video has important practical significance.

In this section, we select several areas in different countries with certain typical surface coverage types, including vegetation, suburbs, and a variety of buildings as the test images. The comparison results on this data set are shown in Fig. 8.

Among them, SRCNN [40] and VDSR [15] produce noticeable sawtooth edges for foreground objects and blurry backgrounds. Comparatively, SRGAN [25] can recover some high-frequency details and sharp image contours, and whereas, it also produces artificial artifacts and dirty effects. By contrast, the proposed EEGAN's results not only show sharp edges but also display clear image content. These comparisons on video satellites further demonstrate the effectiveness of the proposed edge processing strategies, and they also demonstrate the utility and robustness in enhancing images with artifacts and noises.

F. Comparisons With State of the Arts on Digitalglobe

To further demonstrate the efficiency of the proposed EEGAN on real scenes, we conduct another group of comparative experiments on the data sets collected from *Digitalglobe*.² Compared to the above-mentioned simulation experiment, here we directly feed the LR test image into the network without downsampling so it is in line with the real degradation process.

²<http://www.spaceview.com/Satellite/digitalglobe/2014/1209/83.html#main>

TABLE IV
COMPARISON RESULTS OF AG AND NIQE ON DIGITALGLOBE WITH THE SCALE FACTOR OF 4

Algorithms Metrics	Bicubic AG/NIQE	SRCCNN* [41] AG/NIQE	VDSR* [15] AG/NIQE	SRGAN* [25] AG/NIQE	EEGAN (Ours) AG/NIQE
Por Una Cabeza (Beijing)	3.389/7.766	4.173/7.047	4.442/6.556	4.411/6.499	4.508/6.445
Pentagon (US)	5.625/7.383	7.288/6.210	7.583/5.963	7.384/5.649	7.542/5.550
Khalifa Stadium (Qatar)	4.819/6.915	6.506/6.491	6.751/5.801	6.630/5.590	6.825/5.305
Solar Power Plant (Sevilla)	6.454/7.059	7.013/5.922	7.166/5.820	7.365/6.515	7.556/5.638
Convention and Exhibition Center (Yunnan)	3.929/7.454	4.482/6.310	4.574/6.080	4.788/5.964	4.603/5.784
Convention and Exhibition Center (Tibet)	1.910/7.831	2.208/6.936	2.265/7.038	2.476/5.939	2.318/5.898
Burj Dubai (Dubai)	6.425/6.876	8.147/5.846	8.311/5.504	8.806/5.376	8.821/4.999
Ferrari theme Park (UAE)	3.558/6.661	4.736/5.889	4.932/5.613	4.697/4.696	4.899/4.515
Kwai Tsing Container Terminals (Hong Kong)	2.936/7.498	3. 355/6.591	3.421/6.757	3.645/5.206	3.578/5.114
Bird's Nest (Beijing)	3.699/7.132	4.442/6.056	4.405/6.028	4.498/5.570	4.605/5.389
Average	4.274/7.257	5.235/6.329	5.385/6.116	5.470/5.700	5.525/5.463

In this section, we select ten representative scenarios as the test samples (including the Pentagon, Bird's Nest, Burj Dubai, solar power plant, etc.). For convenience, we crop them into a uniform size in 600×456 pixels for testing. We show the visual comparison results in Fig. 9. In the “solar power plant” scene, most of the compared methods produce noticeable artifacts and blurred outlines, whereas the proposed EEGAN yields sharper edges with fewer jagged lines and ringing artifacts. In addition, because of the lacking of HR ground truth, we use two representative indicators without reference (AG and NIQE) for assessment. The comparison results are reported in Table IV. Note that the larger the AG and the smaller the NIQE, the clearer the image. Obviously, in most cases, the proposed EEGAN shows a huge advantage in getting the highest score. In summary, these substantial improvements in visual performance and quantitative indicators in reconstruction performance fully demonstrate the effectiveness and practicability of the EEGAN algorithm in superresolving satellite imagery when combined with the adversarial learning strategy. In addition, the comparison of satellite images in unknown degradation scenarios further demonstrates the superiority of our proposed EEGAN compared to the state-of-the-art methods [15], [25], [40].

V. CONCLUSION

In this paper, we construct a new GAN-based framework for satellite imagery SR reconstruction that learns fine image textures and detailed edges simultaneously. In addition, we propose an edge-enhancement strategy to enforce the image contours, which can effectively remove the artifacts and noises produced by the adversarial training, leading to clean and easily distinguishable edges. Experimentally, we have shown that our method can reconstruct sharp edges and clean image contents, more realistic and faithful to the ground truth. In particular, the proposed EEGAN also exhibits superior competitive performance under unknown degradation conditions.

REFERENCES

- [1] Z. Shao and J. Cai, “Remote sensing image fusion with deep convolutional neural network,” *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 5, pp. 1656–1669, May 2018.
- [2] S. Gou, S. Liu, S. Yang, and L. Jiao, “Remote sensing image super-resolution reconstruction based on nonlocal pairwise dictionaries and double regularization,” *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 12, pp. 4784–4792, Dec. 2014.
- [3] J. P. Duffy, J. D. Shutler, M. J. Witt, L. DeBell, and K. Anderson, “Tracking fine-scale structural changes in coastal dune morphology using kite aerial photography and uncertainty-assessed structure-from-motion photogrammetry,” *Remote Sens.*, vol. 10, no. 9, p. 1494, 2018.
- [4] J. Jiang, J. Ma, C. Chen, Z. Wang, Z. Cai, and L. Wang, “SuperPCA: A superpixelwise PCA approach for unsupervised feature extraction of hyperspectral imagery,” *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 8, pp. 4581–4593, Aug. 2018.
- [5] N. He, L. Fang, S. Li, A. Plaza, and J. Plaza, “Remote sensing scene classification using multilayer stacked covariance pooling,” *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 12, pp. 6899–6910, Dec. 2018.
- [6] L. Fang, G. Liu, S. Li, P. Ghamisi, and J. A. Benediktsson, “Hyperspectral image classification with squeeze multibias network,” *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 3, pp. 1291–1301, Mar. 2019.
- [7] J. Zhu, L. Fang, and P. Ghamisi, “Deformable convolutional neural networks for hyperspectral image classification,” *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 8, pp. 1254–1258, Aug. 201.
- [8] S. Bredemeyer, F.-G. Ulmer, T. H. Hansteen, and T. R. Walter, “Radar path delay effects in volcanic gas plumes: The case of Lásar volcano, northern chile,” *Remote Sens.*, vol. 10, no. 10, p. 1514, 2018.
- [9] C. Li, Y. Ma, X. Mei, C. Liu, and J. Ma, “Hyperspectral unmixing with robust collaborative sparse regression,” *Remote Sens.*, vol. 8, no. 7, p. 588, Jul. 2016.
- [10] X. Liu, D. Zhai, R. Chen, X. Ji, D. Zhao, and W. Gao, “Depth restoration from RGB-D data via joint adaptive regularization and thresholding on manifolds,” *IEEE Trans. Image Process.*, vol. 28, no. 3, pp. 1068–1079, Mar. 2019.
- [11] X. Liu, D. Zhai, R. Chen, X. Ji, D. Zhao, and W. Gao, “Depth super-resolution via joint color-guided internal and external regularizations,” *IEEE Trans. Image Process.*, vol. 28, no. 4, pp. 1636–1645, Apr. 2019.
- [12] F. Li, X. Jia, and D. Fraser, “Universal HMT based super resolution for remote sensing images,” in *Proc. IEEE Conf. ICIP*, Oct. 2008, pp. 333–336.
- [13] M. T. Merino and J. Nunez, “Super-resolution of remotely sensed images with variable-pixel linear reconstruction,” *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 5, pp. 1446–1457, May 2007.
- [14] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in *Proc. IEEE Conf. CVPR*, Jun. 2016, pp. 770–778.
- [15] J. Kim, J. K. Lee, and K. M. Lee, “Accurate image super-resolution using very deep convolutional networks,” in *Proc. CVPR*, Jun. 2016, pp. 1646–1654.
- [16] Y. Luo, L. Zhou, S. Wang, and Z. Wang, “Video satellite imagery super resolution via convolutional neural networks,” *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 12, pp. 2398–2402, Dec. 2017.
- [17] K. Jiang, Z. Wang, P. Yi, and J. Jiang, “A progressively enhanced network for video satellite imagery superresolution,” *IEEE Signal Process. Lett.*, vol. 25, no. 11, pp. 1630–1634, Nov. 2018.
- [18] T. Tong, G. Li, X. Liu, and Q. Gao, “Image super-resolution using dense skip connections,” in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 4809–4817.
- [19] J. Pan *et al.*, “Learning dual convolutional neural networks for low-level vision,” in *Proc. IEEE Conf. CVPR*, Jun. 2018, pp. 3073–3079.
- [20] K. Jiang, Z. Wang, P. Yi, J. Jiang, J. Xiao, and Y. Yao, “Deep distillation recursive network for remote sensing imagery super-resolution,” *Remote Sens.*, vol. 10, no. 11, p. 1700, 2018.

- [21] Z. Wang *et al.*, "Multi-memory convolutional neural network for video super-resolution," *IEEE Trans. Image Process.*, vol. 28, no. 5, pp. 2530–2544, May 2018.
- [22] I. Goodfellow *et al.*, "Generative adversarial nets," in *Proc. Int. Conf. NIPS*, 2014, pp. 2672–2680.
- [23] M. S. M. Sajjadi, B. Schölkopf, and M. Hirsch, "EnhanceNet: Single image super-resolution through automated texture synthesis," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 4501–4510.
- [24] X. Zhang, X. Zhu, X.-Y. Zhang, N. Zhang, P. Li, and L. Wang, "SegGAN: Semantic segmentation with generative adversarial network," in *Proc. IEEE 4th Int. Conf. BigMM*, Sep. 2018, pp. 1–5.
- [25] C. Ledig *et al.*, "Photo-realistic single image super-resolution using a generative adversarial network," in *Proc. IEEE Conf. CVPR*, Jul. 2017, pp. 4681–4690.
- [26] J. Jiang, C. Chen, Y. Yu, X. Jiang, and J. Ma, "Spatial-aware collaborative representation for hyperspectral remote sensing image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 3, pp. 404–408, Mar. 2017.
- [27] F. Tong, H. Tong, J. Jiang, and Y. Zhang, "Multiscale union regions adaptive sparse representation for hyperspectral image classification," *Remote Sens.*, vol. 9, no. 9, p. 872, 2017.
- [28] J. Jiang, J. Ma, Z. Wang, C. Chen, and X. Liu, "Hyperspectral image classification in the presence of noisy labels," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 2, pp. 851–865, Feb. 2018.
- [29] Q. Mao, S. Wang, S. Wang, X. Zhang, and S. Ma, "Enhanced image decoding via edge-preserving generative adversarial networks," in *Proc. IEEE Int. Conf. ICME*, Jul. 2018, pp. 1–6.
- [30] W. Yang *et al.*, "Deep edge guided recurrent residual learning for image super-resolution," *IEEE Trans. Image Process.*, vol. 26, no. 12, pp. 5895–5907, Dec. 2017.
- [31] W. Shi *et al.*, "Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 1874–1883.
- [32] B. Kamgar-Parsi, B. Kamgar-Parsi, and A. Rosenfeld, "Optimally isotropic Laplacian operator," *IEEE Trans. Image Process.*, vol. 8, no. 10, pp. 1467–1472, Oct. 1999.
- [33] W.-S. Lai, J.-B. Huang, N. Ahuja, and M.-H. Yang, "Deep laplacian pyramid networks for fast and accurate super-resolution," in *Proc. Int. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jul. 2017, pp. 5835–5843.
- [34] Y. Tai, J. Yang, and X. Liu, "Image super-resolution via deep recursive residual network," in *Proc. IEEE Conf. CVPR*, Jul. 2017, pp. 3147–3155.
- [35] B. Lim, S. Son, H. Kim, S. Nah, and K. M. Lee, "Enhanced deep residual networks for single image super-resolution," in *Proc. Comput. Vis. Pattern Recognit. Workshops*, Jul. 2017, pp. 1132–1140.
- [36] L. Zhou, Z. Wang, Y. Luo, and Z. Xiong, "Separability and compactness network for image recognition and superresolution," *IEEE Trans. Neural Netw. Learn. Syst.*, to be published.
- [37] W. Dong *et al.*, "Hyperspectral image super-resolution via non-negative structured sparse representation," *IEEE Trans. Image Process.*, vol. 25, no. 5, pp. 2337–2352, May 2016.
- [38] S. Yang, F. Sun, M. Wang, Z. Liu, and L. Jiao, "Novel super resolution restoration of remote sensing images based on compressive sensing and example patches-aided dictionary learning," in *Proc. Int. Workshop Multi-Platform/Multi-Sensor Remote Sens. Mapping*, Jan. 2011, pp. 1–6.
- [39] H. Chavez-Roman and V. Ponomaryov, "Super resolution image generation using wavelet domain interpolation with edge extraction via a sparse representation," *IEEE Geosci. Remote Sens. Lett.*, vol. 11, no. 10, pp. 1777–1781, Oct. 2014.
- [40] 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.
- [41] C. Dong, C. C. Loy, and X. Tang, "Accelerating the super-resolution convolutional neural network," in *Proc. 14th Eur. Conf. Comput. Vis. (ECCV)*, Oct. 2016, pp. 391–407.
- [42] 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.
- [43] M. Zhang, M. Gong, Y. Mao, J. Li, and Y. Wu, "Unsupervised feature extraction in hyperspectral images based on wasserstein generative adversarial network," *IEEE Trans. Geosci. Remote Sens.*, to be published.
- [44] J. Johnson, A. Alahi, and L. Fei-Fei, "Perceptual losses for real-time style transfer and super-resolution," in *Proc. IEEE Conf. ECCV*, B. Leibe, J. Matas, N. Sebe, and M. Welling, Eds. Cham, Switzerland: Springer, 2016, pp. 694–711.
- [45] C. K. Sønderby, J. Caballero, L. Theis, W. Shi, and F. Huszár, (Oct. 2016). "Amortised MAP inference for image super-resolution." [Online]. Available: <https://arxiv.org/abs/1610.04490>
- [46] B. Wu, H. Duan, Z. Liu, and G. Sun, (Dec. 2017). "SRPGAN: Perceptual generative adversarial network for single image super resolution." [Online]. Available: <https://arxiv.org/abs/1712.05927>
- [47] L. Xu, J. Ren, Q. Yan, R. Liao, and J. Jia, "Deep edge-aware filters," in *Proc. Int. Conf. ICML*, 2015, pp. 1669–1678.
- [48] S. Liu, J. Pan, and M.-H. Yang, "Learning recursive filters for low-level vision via a hybrid neural network," in *Proc. IEEE Int. Conf. ECCV*. Cham, Switzerland: Springer, 2016, pp. 560–576.
- [49] M. T. Alonso, C. López-Martínez, J. J. Mallorquí, and P. Salembier, "Edge enhancement algorithm based on the wavelet transform for automatic edge detection in SAR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 1, pp. 222–235, Jan. 2011.
- [50] P. Bao, L. Zhang, and X. Wu, "Canny edge detection enhancement by scale multiplication," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 9, pp. 1485–1490, Sep. 2005.
- [51] G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE Conf. CVPR*, Jul. 2017, pp. 2261–2269.
- [52] G. Huang, D. Chen, T. Li, F. Wu, L. van der Maaten, and K. Q. Weinberger, (Mar. 2018). "Multi-scale dense networks for resource efficient image classification." [Online]. Available: <https://arxiv.org/abs/1703.09844>
- [53] Y. Tai, J. Yang, X. Liu, and C. Xu, "MemNet: A persistent memory network for image restoration," in *Proc. IEEE Int. Conf. ICCV*, Oct. 2017, pp. 4539–4547.
- [54] R. K. Srivastava, K. Greff, and J. Schmidhuber, "Training very deep networks," in *Proc. Adv. NIPS*, 2015, pp. 2377–2385.
- [55] J. Kim, J. Lee, and K. M. Lee, (2015). "Accurate image super-resolution using very deep convolutional networks." [Online]. Available: <https://arxiv.org/abs/1511.04587>
- [56] Y. Tai, J. Yang, X. Liu, and C. Xu, "MemNet: A persistent memory network for image restoration," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2017, pp. 4549–4557.
- [57] X. Wang, "Laplacian operator-based edge detectors," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 5, pp. 886–890, May 2007.
- [58] C. Xiansheng, "An edge detection new algorithm based on laplacian operator," in *Proc. IEEE 3rd Int. Conf. Commun. Softw. Netw.*, May 2011, pp. 202–206.
- [59] J. Kim, J. K. Lee, and K. M. Lee, "Deeply-recursive convolutional network for image super-resolution," in *Proc. IEEE Conf. CVPR*, Jun. 2016, pp. 1637–1645.
- [60] L. Zhang, L. Zhang, X. Mou, and D. Zhang, "FSIM: A feature similarity index for image quality assessment," *IEEE Trans. Image Process.*, vol. 20, no. 8, pp. 2378–2386, Aug. 2011.
- [61] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.
- [62] A. Chen, B. Chen, X. Chai, R. Bian, and H. Li, (2017). "A novel stochastic stratified average gradient method: Convergence rate and its complexity." [Online]. Available: <https://arxiv.org/abs/1710.07783>
- [63] A. Mittal, R. Soundararajan, and A. C. Bovik, "Making a 'completely blind' image quality analyzer," *IEEE Signal Process. Lett.*, vol. 20, no. 3, pp. 209–212, Mar. 2013.



Kui Jiang received the B.S. degree from the College of Chemistry and Chemical Engineering, Xinjiang University, Ürümqi, China, in 2017. He is currently pursuing the Ph.D. degree with the School of Computer, Wuhan University, Wuhan, China, under the supervision of Prof. Z. Wang.

His research interests include image/video processing and computer vision.



Zhongyuan Wang (M'13) received the Ph.D. degree in communication and information system from Wuhan University, Wuhan, China, in 2008.

He is currently an Associate Professor with the School of Computer, Wuhan University. He is currently directing three projects funded by the National Natural Science Foundation Program of China. His research interests include video compression, image processing, and multimedia communications.



Peng Yi received the B.S. degree from the Faculty of Electronic Information and Electrical Engineering, Dalian University of Technology, Dalian, China, in 2017. He is currently pursuing the Ph.D. degree with the School of Computer, Wuhan University, Wuhan, China, under the supervision of Prof. Z. Wang.

His research interests include image/video processing and computer vision.



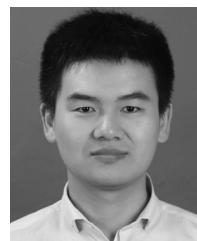
Tao Lu received the B.S and M.S degrees from the School of Computer Science and Engineering, Wuhan Institute of Technology, Wuhan, China, in 2003 and 2008, respectively, and the Ph.D. degree from the National Engineering Research Center For Multimedia Software, Wuhan University, in 2013.

From 2015 to 2017, he was a Post-Doctoral Researcher with the Department of Electrical and Computer Engineering, Texas A&M University, College Station, TX, USA. He is currently an Associate Professor with the School of Computer Science and Engineering, Wuhan Institute of Technology, where he is also a Research Member with the Hubei Provincial Key Laboratory of Intelligent Robot. His research interests include image/video processing, computer vision, and artificial intelligence.



Guangcheng Wang received the B.S. degree from the School of Electrical and Optoelectronic Engineering, Changzhou Institute of Technology, Changzhou, China, in 2015, and the M.S. degree from the School of Information and Control Engineering, China University of Mining and Technology, Xuzhou, China, in 2018. He is currently pursuing the Ph.D. degree with the National Engineering Research Center for Multimedia Software, School of Computer, Wuhan University, Wuhan, China.

His research interests include image processing and quality assessment.



Junjun Jiang (M'15) received the B.S. degree from the Department of Mathematics, Huaqiao University, Quanzhou, China, in 2009, and the Ph.D. degree from the School of Computer, Wuhan University, Wuhan, China, in 2014.

From 2015 to 2018, he was an Associate Professor with the China University of Geosciences, Wuhan. From 2016 to 2018, he was a Project Researcher with the National Institute of Informatics, Tokyo, Japan. He is currently a Professor with the School of Computer Science and Technology, Harbin Institute of Technology, Harbin, China. His research interests include image processing and computer vision.

Dr. Jiang was a recipient of the World's FIRST 10K Best Paper Award (Finalist) at ICME 2017, the Best Student Paper Runner-up Award at MMM 2015, the Best Paper Award at IFTC 2018, the 2016 China Computer Federation Outstanding Doctoral Dissertation Award, and the 2015 ACM Wuhan Doctoral Dissertation Award.