SUPER-RESOLUTION OF SATELLITE IMAGES BY TWO-DIMENSIONAL RRDB AND EDGE-ENHANCEMENT GENERATIVE ADVERSARIAL NETWORK

Yu-Zhang Chen[⋆] *Tsung-Jung Liu*[⋆] *Kuan-Hsien Liu*[†]

[⋆]Department of Electrical Engineering and Graduate Institute of Communication Engineering, National Chung Hsing University, Taiwan

[†]Dept. of Computer Science and Information Engineering, National Taichung University of Science and Technology, Taiwan

ABSTRACT

With the increasing demand for high-resolution images, image super-resolution (SR) technology has become one of the focuses in related research fields. Generally speaking, high resolution is usually achieved by increasing the density and accuracy of the sensor. However, such an approach is quite expensive for equipment and design. In particular, increasing the density of satellite sensors must be undertaken great risks. Inspired by EEGAN and based on it, the Ultra-Dense Subnet (UDSN) and Edge Enhanced Network (EEN) were modified. Among them, the UDSN is used for feature extraction and obtains high-resolution results that look clear in the intermediate but are deteriorated by artifacts, and the Edge-Enhanced Subnet (EESN) is used to purify, extract and enhance the image contour and use mask processing to eliminate images contaminated by noise. Finally, the restored intermediate image and the enhanced edge are combined to produce a high-resolution image with high credibility and clear content. We use Kaggle and AID open experimental datasets to test and compare the results among different methods. It proves the performance of the proposed model is better than other SR methods.

Index Terms— Super-resolution(SR), satellite images, generative adversarial network (GAN), residual in residual dense block (RRDB)

1. INTRODUCTION

In recent years, the application of high-resolution (HR) satellite imagery is crucial in urban planning, environmental disaster assessment, building extraction, and small object detection. However, due to the limitations of the underlying technology and the high cost of the hardware, the observed HR images often have incomplete spatial and temporal coverage, and the resolution does not meet the requirements, making them unable to meet the growing production and applications. This has a negative impact on the accuracy of subsequent computer vision tasks. Image super-resolution (SR) technology provides a low-cost and effective method of acquiring HR images by reconstructing HR images from relatively low-resolution (LR) but easily available images. Therefore, this paper mainly discusses how to provide high-quality satellite images in a cost-effective manner.

Recently, the methods based on Generative Adversarial Networks (GAN) [1], such as Super Resolution GAN (SRGAN) [2] and Enhanced Super Resolution GAN (ESRGAN) [3], have outstanding performance in enhancing low-resolution images with or without noise. These models consist of two sub-networks: generators and discriminators. Both sub-networks are composed of deep CNN. A dataset containing HR and LR image pairs is used to train and test the model. The generator generates an HR image from the LR input image, and the discriminator predicts whether the generated image is a real HR image or an enlarged LR image. After sufficient training, the generator generates HR images similar to real HR images, and the discriminator can no longer correctly distinguish between true and false images.

This paper is inspired by the EEGAN [4] and EESRGAN [5] networks and uses EEGAN as the model infrastructure to propose a feasible model architecture and training process. For the generator and edge enhancement network, a combination of two-dimensional topology and residual in residual dense block (2D-RRDB) is used. Compared with the conventional residual in residual dense block (RRDB) [3], this two-dimensional structure with additional diagonal lines can provide more possibilities for information conversion and bring better gradient optimization on the connection between different paths. On the other hand, through the diagonal connection method, more sufficient link paths can be obtained under the same number of layers. In this way, it can effectively improve the training difficulty, gradient disappearance, information propagation disappearance problems caused by increasing the depth of the layer for increasing the density of dense block connections under the traditional one-dimensional structure.

In addition, in view of the use of the same method to calculate the perceptual loss of the entire image (i.e., this means that the same features are used on the edges, foreground or background), this method enables the network to consider unnecessary losses and learn the features with smaller amount of information. (for example: the texture of the building). We use the feature map before the activation layer of VGG19 [6] to calculate the perceptual loss function and edge percep-

tual loss. This improvement will provide sharper edges and more visually consistent results. In addition, considering that satellite images often bring more noise than ordinary images, the Canny operator method [7] is introduced for edge extraction to help the generator generate clearer and distinguishable edge information. Finally, we used open experimental data set for training and testing, and compared the results with other SR methods. We not only consider PSNR to evaluate our method, but also look for several evaluation methods. As a result, we learned that our method can obtain satellite images that are more natural and visually closer to real images.

2. METHOD

In this section, we introduce the proposed method in detail. We use EEGAN [4] as the basic architecture and focus on modification. The generator G can be roughly divided into two sub-structures: ultra-dense subnet (UDSN) and edge-enhanced subnet (EESN). UDSN is composed of multiple dense blocks and a reconstruction layer used to generate intermediate high-resolution image results. EESN is used to enhance the target edge extracted from the intermediate super-resolution image by removing noise and artifacts. Finally, the noisy edge in the middle SR image is replaced by the purified edge from EESN to obtain the final SR output. We also tried to add a perceptual loss function that can improve visual quality [8, 9, 10]. Through the adjustment of the parameters, we find a set of better parameters that can obtain higher performance. We also compare the obtained high-resolution images with the ones generated by other methods. More details will be described in subsequent sections.

2.1. Generator Network

As shown in Figure 1, we first modify the UDSN used to generate intermediate high-resolution images, and then modify the dense subnet branch used for the EESN. For the UDSN, we replace the original dense block with the RRDB, and use a convolutional layer with a two-dimensional topology, as shown in Figure 2. Through our experimental results, the improvement of this network architecture helps the generation network to generate SR images that are closer to real images. Then we use a two-dimensional topological convolutional layer to replace this RRDB in the dense network branch of the EESN. We hope that through this network architecture, the EESN can generate better edge features.

2.2. Edge Extraction and Enhancement

In this subsection, we mainly discuss the feature extraction methods used in the EESN of the generator network. We use different edge extraction methods and compare their effects through experiments to find better methods. In addition, we also quoted the Mask Branch in EEGAN [4] to suppress noise

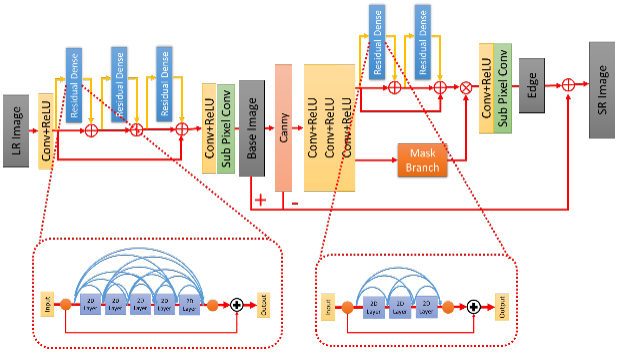

Fig. 1. Proposed network architecture

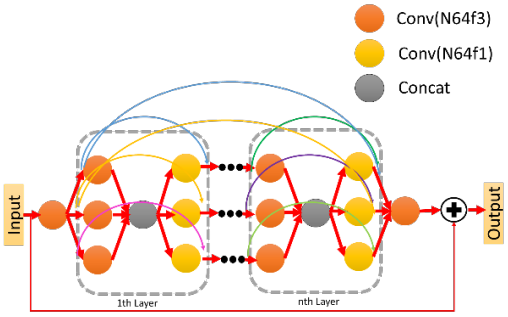

Fig. 2. Convolutional layer with a two-dimensional topology

and false edges. First, the intermediate SR image generated by the previous UDSN is used as input, and a basic edge image is obtained through the edge extraction operator, and then this edge image is thrown to the two subnetworks (i.e., edge extraction and mask branches) separately. The branches eventually produce sharper edges. A more detailed operation is shown in Figure 3. In the experiment, we used the Sobel and Canny methods [11]. As a result, we found that Canny has a better effect on the final SR image.

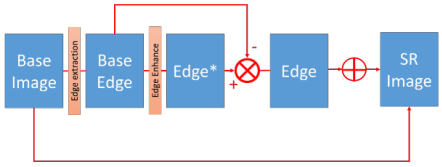

Fig. 3. Edge extraction and enhancement operation method

2.3. Loss Function

We first build a content loss function to force the generator G to generate an intermediate HR image similar to the real HR image using the following model:

$$Loss_{cont}(\tau_G) = arg\min_{\tau_G} \sum_{i=1}^{k} \rho(I_{HR} - I_{BASE}), \quad (1)$$

where τ_G represents a set of model parameters in the generator, $\rho(x) = \sqrt{x^2 + \varepsilon^2}$ represents the Charbonnier penalty function [12]. We set the compensation parameter to $\varepsilon = 10^{-3}$, I_{HR} and I_{BASE} refer to the real HR image and the intermediate HR image generated by UDSN.

In order to improve the quality of the reconstructed image and reduce artifacts, we added the pixel-based Charbonnier loss to enhance the consistency of the image content between the generated HR image and the real HR image. The consistency loss function is:

$$Loss_{img\_cst}(\tau_G) = \rho(I_{HR} - I_{SR}), \tag{2}$$

where τ_G represents a set of model parameters in the generator, I_{HR} and I_{SR} refer to the real HR image and the final HR image.

We input I_{HR} and I_{BASE} into the discriminator to determine their authenticity. We train the discriminator to minimize the adversarial loss, which can encourage the generator to generate a reconstructed image I_{SR} close to the real image, which is defined as:

$$Loss_{adv}(\tau_G, \tau_D) = -\log D(I_{HR}) - \log(1 - D(G(I_{LR}))), \tag{3}$$

where τ_D refers to the model parameters in the discriminator, I_{LR} represents the low-resolution image input, $G(\cdot)$ represents the generating function network, and $D(\cdot)$ is the discrimination function to calculate the probability of being the real HR image or generated HR image. For the discriminator, the generated image wants to be classified as 0, and the real image wants to be classified as 1. If the discriminator can correctly distinguish real or generated, then formula (3) will get the smallest loss 0. For the generator, we hope to confuse the discriminator so that it mistakes the generated image for a real image and classifies the image as 1.

Then we added the image edge consistency loss function, as shown in formula (4), where I_{edge_HR} is the edge image of the real HR image, and I_{edge_SR} is the edge image of the generated SR image. Although the image consistency loss $Loss_{img\_cst}$ in formula (2) helps to obtain an output with good edge information, the edges of some objects sometimes are distorted and produce some noise. Thus we add this loss to solve this problem. We give it the same weight as $Loss_{img\_cst}$.

$$Loss_{edge\_cst}(\tau_G) = \rho(I_{edge\_HR} - I_{edge\_SR}). \tag{4}$$

The perceptual loss [13] is shown in formula (5), and we also added the edge perceptual loss function shown in formula (6). We use the feature map ($vgg_{fea}(\cdot)$) before the activation layer of the fine-tuned VGG19 [6] network to calculate the perceptual loss and the perceptual edge loss.

$$L_{perc} = E\left\|vgg_{fea}(I_{BASE}) - vgg_{fea}(I_{HR})\right\|, \tag{5}$$

$$L_{perc-edge} = E\left\|vgg_{fea}(I_{edge\_BASE}) - vgg_{fea}(I_{edge\_HR})\right\|, \tag{6}$$

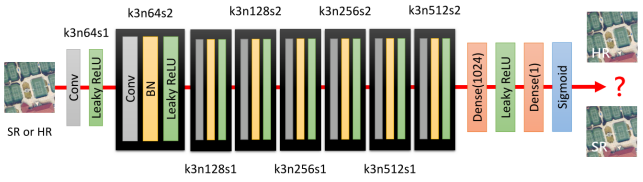

Fig. 4. Network architecture of the discriminator

where I_{BASE} represents the intermediate generated HR image, I_{HR} is the real HR image, I_{edge_BASE} is the edge of the intermediate generated image, I_{edge_HR} is the edge of the real image, and E represents the image mean calculation.

The final loss function is formula (8), in which we combine $Loss_{img\_cst}(\tau_G)$ and $Loss_{edge\_cst}(\tau_G)$ into $Loss_{cst}(\tau_G)$, as shown in formula (7). According to the previous experimental experience, we set $\alpha_1 = 0.1$, $\alpha_2 = 0.009$.

$$Loss_{cst}(\tau_G) = Loss_{img\_cst}(\tau_G) + Loss_{edge\_cst}(\tau_G), \tag{7}$$

$$\begin{aligned} Loss_{final}(\tau_G, \tau_D) = Loss_{cont}(\tau_G) + 0.001 Loss_{adv}(\tau_G, \tau_D) \\ +5 Loss_{cst}(\tau_G) + \alpha_1 L_{perc} + \alpha_2 L_{perc-edge}. \end{aligned} \tag{8}$$

2.4. Discriminator Network

We use the standard generative adversarial network to evaluate the probability that the image is true and natural. In order to distinguish which images are generated images and which are real image samples, we trained a discriminator model. The network structure of the discriminator is shown in Figure 4. We use LeakyReLU activation ($\alpha = 0.2$) and use stride convolution instead of the maximum pooling layer. The purpose of stride convolution is to reduce the size of the feature. The network contains eight convolutional layers, except for the first convolutional layer, all other convolutional layers are followed by a batch normalization layer (BN). At the end of the network, there are two dense layers and a sigmoid activation function which can compute the probability of image authenticity.

3. EXPERIMENTS

We compare our method with several satellite image super-resolution methods based on generative adversarial networks, including SRGAN [2], ESRGAN [3], EEGAN [4], and EESRGAN [5]. The test was conducted on a public satellite image database used by most researchers. And we also used several commonly used objective evaluation indicators [14, 15] (namely peak signal-to-noise ratio (PSNR), feature similarity (FSIM) [16] and structural similarity (SSIM) [17]) to test the performance.

Table 1. Kaggle dataset result comparison

	SRGAN	ESRGAN	EEGAN	EESRGAN	Ours
PSNR	32.215	32.396	32.463	32.804	**33.001**
SSIM	0.893	0.887	0.901	0.899	**0.906**
FSIM	0.987	0.992	0.991	0.990	**0.993**

Table 2. AID dataset result comparison

	SRGAN	ESRGAN	EEGAN	EESRGAN	Ours
PSNR	32.164	32.250	32.140	32.335	**32.399**
SSIM	0.876	0.883	0.903	0.896	**0.908**
FSIM	0.980	0.989	0.990	**0.992**	0.990

3.1. Databases and Experimental Parameter Setting

Kaggle open source data set [18] contains 1,720 satellite images with a size of 3,099×2,329 pixels. In the training process of this data set, first of all, in order to increase the amount of data, we crop the image into 720×720. The data set has a total of 20,640 images, and we use 80% of them as the training set and 20% as the test set. Then we train the input low-resolution image with a size of 180×180. For the AID data set [19], it has a size of 600×600 with a total of 10,000 images and is composed of 30 scene types. In the training process of this data set, we also use 80% of them as the training set and 20% as the test set. During the training process, we use Adam to optimize and set the parameter β_1 to 0.9. The learning rate is initialized to 2×10^{-4}. We alternately train the generator and discriminator until the model converges.

3.2. Comparisons with State-of-the-Art Methods on Kaggle and AID Open Source Data Set

To illustrate the practicality, we compare our method with several GAN-based SR methods, including SRGAN [2], ESRGAN [3], EEGAN [4], and EESRGAN [5]. For a fair comparison, we retrain these models using the same training data set. As shown in Table 1, our method shows the highest scores for all indicators in Kaggle dataset, including PSNR, SSIM and FSIM. Also, in Figure 5, we can observe the results of ours are still very competitive and more realistic compared with other methods. We also compare the proposed model with state-of-the-art methods in the AID data set. As shown in Table 2, our method still has good scores in various indicators. Figure 6 shows the visual result comparison for all compared methods.

Fig. 5. Kaggle dataset visual result comparison

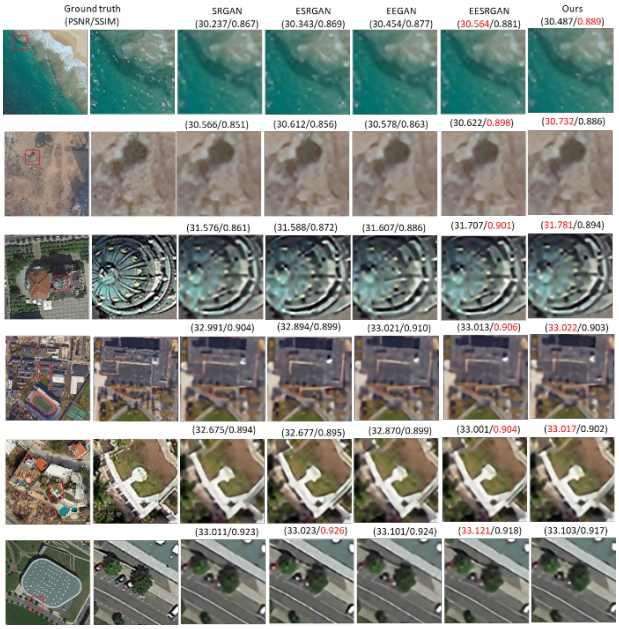

Fig. 6. AID dataset visual result comparison

4. CONCLUSION

In this paper, we combine the residual dense block and the convolutional network with two-dimensional topology to form a network architecture with complex connections. In the generation sub-network, it can provide the generation of real images closer to the ground truth. For images, better edge features can be extracted in the EESN. In addition, we have replaced the feature extraction method. And we add different loss functions to improve the quality of the resulting image. Through experiments, our super-resolution method can produce images with better visual quality [20]. According to the results of objective evaluation [21], our proposed method is better than the current satellite imagery SR methods.

5. REFERENCES

[1] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio, "Generative adversarial nets," *Advances in neural information processing systems*, vol. 27, 2014.

[2] Christian Ledig, Lucas Theis, Ferenc Huszár, Jose Caballero, Andrew Cunningham, Alejandro Acosta, Andrew Aitken, Alykhan Tejani, Johannes Totz, Zehan Wang, et al., "Photo-realistic single image super-resolution using a generative adversarial network," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 4681–4690.

[3] Xintao Wang, Ke Yu, Shixiang Wu, Jinjin Gu, Yihao Liu, Chao Dong, Yu Qiao, and Chen Change Loy, "Esrgan: Enhanced super-resolution generative adversarial networks," in *Proceedings of the European conference on computer vision (ECCV) workshops*, 2018, pp. 0–0.

[4] Kui Jiang, Zhongyuan Wang, Peng Yi, Guangcheng Wang, Tao Lu, and Junjun Jiang, "Edge-enhanced gan for remote sensing image superresolution," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 8, pp. 5799–5812, 2019.

[5] Jakaria Rabbi, Nilanjan Ray, Matthias Schubert, Subir Chowdhury, and Dennis Chao, "Small-object detection in remote sensing images with end-to-end edge-enhanced gan and object detector network," *Remote Sensing*, vol. 12, no. 9, pp. 1432, 2020.

[6] Karen Simonyan and Andrew Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[7] John Canny, "A computational approach to edge detection," *IEEE Transactions on pattern analysis and machine intelligence*, , no. 6, pp. 679–698, 1986.

[8] Tsung-Jung Liu, Weisi Lin, and C-C Jay Kuo, "Image quality assessment using multi-method fusion," *IEEE Transactions on image processing*, vol. 22, no. 5, pp. 1793–1807, 2012.

[9] Tsung-Jung Liu, Kuan-Hsien Liu, Joe Yuchieh Lin, Weisi Lin, and C-C Jay Kuo, "A paraboost method to image quality assessment," *IEEE transactions on neural networks and learning systems*, vol. 28, no. 1, pp. 107–121, 2015.

[10] Tsung-Jung Liu and Kuan-Hsien Liu, "No-reference image quality assessment by wide-perceptual-domain scorer ensemble method," *IEEE Transactions on Image Processing*, vol. 27, no. 3, pp. 1138–1151, 2017.

[11] GT Shrivakshan and Chandramouli Chandrasekar, "A comparison of various edge detection techniques used in image processing," *International Journal of Computer Science Issues (IJCSI)*, vol. 9, no. 5, pp. 269, 2012.

[12] Wei-Sheng Lai, Jia-Bin Huang, Narendra Ahuja, and Ming-Hsuan Yang, "Deep laplacian pyramid networks for fast and accurate super-resolution," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 624–632.

[13] Justin Johnson, Alexandre Alahi, and Li Fei-Fei, "Perceptual losses for real-time style transfer and super-resolution," in *European conference on computer vision*. Springer, 2016, pp. 694–711.

[14] Tsung-Jung Liu, Hsin-Hua Liu, Soo-Chang Pei, and Kuan-Hsien Liu, "A high-definition diversity-scene database for image quality assessment," *IEEE Access*, vol. 6, pp. 45427–45438, 2018.

[15] Tsung-Jung Liu, "Study of visual quality assessment on pattern images: Subjective evaluation and visual saliency effects," *IEEE Access*, vol. 6, pp. 61432–61444, 2018.

[16] Lin Zhang, Lei Zhang, Xuanqin Mou, and David Zhang, "Fsim: A feature similarity index for image quality assessment," *IEEE transactions on Image Processing*, vol. 20, no. 8, pp. 2378–2386, 2011.

[17] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE transactions on image processing*, vol. 13, no. 4, pp. 600–612, 2004.

[18] "Kaggle open source data set. [Online]. Available: https://www.kaggle.com/c/draper-satellite-image-chronology/data," 2016.

[19] Gui-Song Xia, Jingwen Hu, Fan Hu, Baoguang Shi, Xiang Bai, Yanfei Zhong, Liangpei Zhang, and Xiaoqiang Lu, "Aid: A benchmark data set for performance evaluation of aerial scene classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 55, no. 7, pp. 3965–3981, 2017.

[20] Tsung-Jung Liu, Weisi Lin, and C-C Jay Kuo, "Recent developments and future trends in visual quality assessment," *APSIPA ASC*, 2011.

[21] Bo-Xun Chen, Tsung-Jung Liu, Kuan-Hsien Liu, Hsin-Hua Liu, and Soo-Chang Pei, "Image super-resolution using complex dense block on generative adversarial networks," in *2019 IEEE International Conference on Image Processing (ICIP)*. IEEE, 2019, pp. 2866–2870.