SUPER RESOLUTION FOR SINGLE SATELLITE IMAGE USING A GENERATIVE ADVERSARIAL NETWORK

Ran Li[1]*, Wangzeng Liu[1], Wenyu Gong[2], Xiuli Zhu[1], Xinpeng Wang[1]

[1]National Geomatics Center of China
[2]State Key Laboratory of Earthquake Dynamics, Institute of Geology, China Earthquake Administration

Commission III, WG IV/b

KEY WORDS: Super Resolution, Satellite Imagery, Generative Adversarial Network, Residual Network

ABSTRACT:

Inspired by the immense success of deep neural network in image processing and object recognition, learning-based image super resolution (SR) methods have been highly valued and have become the mainstream direction of super resolution research. Base on the recent proposed state-of-art convolution neural network (CNN) super-resolution methods, this paper proposed a generative adversarial network for single satellite image Super Resolution reconstruction. It built on a trained deep residual network to generate preliminary SR images, combined with a discriminative network learns to differentiate preliminary SR images and High resolution samples. The experiments results show that our method can use existing model parameters to refine SR image performance.

1. INTRODUCTION

The need for very high resolution images has grown in various applications, including environmental investigation and monitoring, urban planning, disaster emergency management, and military applications (Benediktsson et al., 2013). However, the satellite resolution generally limited by the spaceborne imaging equipment and orbital altitude. In addition, satellites are effected by communication bandwidth, atmospheric turbulence, transmission noise, and motion blur. The quality and resolution of images from remote sensing satellites cannot meet the requirement of the growing needs. Hence, super resolution (SR) technology, which can reconstruct from lower resolution images and improve the spatial resolution through algorithm manner, is particularly urgent.

Previously, SR methods in remote sensing satellites images rely on data fusion technology with images obtained by different sensor, such as Landsat images by fusing multi-bands image to complementary information (tsai, R.Y., 1984). Another well-known example is SPOT-5, which reaches 2.5m resolution through the SR of two 5m images sampled from shifting a double Charge-Couple Device (CCD) array by subpixel sampling interval (Lim et al., 2009; Nasrollahi et al., 2014). Motivated by recent successes achieved with deep neural network in image processing and object recognition, Lu et al. (Lu et al., 2019) presented a multi-scale residual neural network (MRNN) that adopts the multi-scale nature of satellites images to accurately reconstruct high-frequency information. Luo et al. (Luo et al., 2017) proposed a robust Convolutional Neural Network (CNN) SR mothed to reconstruct high-resolution video satellite images. However, satellites images reconstruction along with the CNN architectures needs massive samples to train millions weights of model parameters (Yu et al., 2017). It is not a general solution for various application purposes with different satellite imagery. To provide a more general solution, this study proposes a generative adversarial network (GAN) with the single satellites image to achieve super resolution. This method employs a trained deep residual network to generate preliminary SR images, combined with a discriminative network learns to differentiate preliminary SR images and high resolution samples. We perform real data experiments on GF2 satellite image to investigate the effectiveness of our method.

* Corresponding author.

2. METHOD

Most state-of-the-art SR algorithm are learning based, which intends to search a nonlinear mapping between I^{LR} and I^{HR}. Thus, our SR method is to train a generating function G that estimates for a given LR (Low Resolution) input image and its HR (High Resolution) counterpart. It is supposed to achieve a better representation for mapping relation between I^{LR} and I^{HR}.

2.1 generative adversarial network architecture

The goal of super resolution is to estimate a high-resolution, super resolution image I^{SR} from low resolution input image I^{LR}. Here I^{LR} is the low-resolution version of its high-resolution counterpart. For an image with C channels, we describe I^{LR} by a byte-valued matrix of size W×H×C, I^{SR} by rW×rH×C respectively, where r is a down sampling factor. First, we employ a trained generator network as a feed-forward residual network $G_{\theta G}$ parametrized by G_G which is obtained by optimizing a SR-specific loss function l^{SR}. For n HR image samples, the generator network can be described in (1).

$$\hat{\theta}_G = argmin\frac{1}{N}\sum_{n=1}^{N} l^{SR}(G_{\theta G}(I_n^{LR}), I_n^{HR}) \qquad (1)$$

Secondly, we define a discriminator network $D_{\theta D}$ which can be optimized by using an alternating manner along with $G_{\theta G}$ to solve the adversarial min-max problem:

$$\min_{\theta G}\max_{\theta D} \quad \|[I^{HR} - p(I^{HR})]logD_{\theta D}(I^{HR})\| + \|[I^{LR} - p(I^{LR})]\log(1 - logD_{\theta D}(G_{\theta G}(I^{LR})))\| \qquad (2)$$

Where $p(I^{HR})$ and $p(I^{LR})$ means pervious train step predict result.

The basic idea of our study is to build a framework for generating realistic looking satellite image based on the adversarial network. It allows us to train a discriminative network D on the purpose of learning differentiate generated SR image from real HR image. With this approach the generator network G can be trained to create a SR image that are highly similar to real HR image until the discriminator network cannot identify. In Eq. (2) SR should be obtained by minimizing the Euclidean distance

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-V-3-2022-591-2022 | © Author(s) 2022. CC BY 4.0 License.

loss between the reconstructed image and its corresponding ground truth high-resolution image.

As illustrated in Fig 1(a), the generator network G is residual blocks with identical layout. (Lim et al. 2018). Specially, it has two convolutional layers with 3 × 3 kernels, which can be modified according the down sampling factor, and 64 feature maps followed by batch-normalization layers and ReLU as activation layer. To discriminate real HR image from generated SR image, we also need to train a discriminator network which is shown in Fig 1b. We follow the architectural proposed by Radford et al. (Radford et al., 2015) and use ReLU activation layer to avoid max-pooling throughout the network.

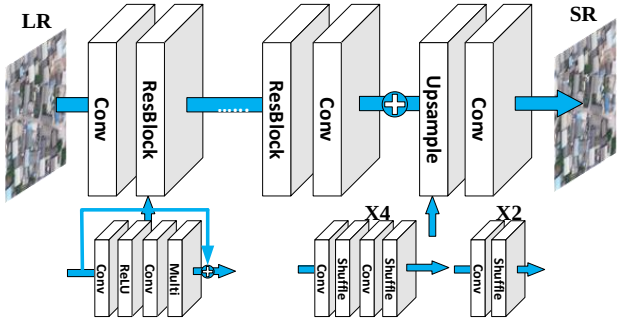

Figure 1 (a). Architecture of Generator Network

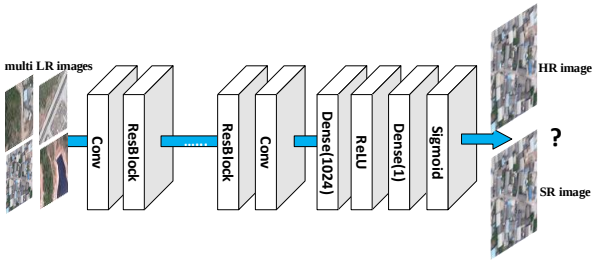

Figure 1 (b). Architecture of Discriminator Network

2.2 Loss function

The definition of the loss function is critical for the SR method performance. Most learning based SR method are built on themean squre error (MSE) (Dong C. et al., 2015), we designed a loss function as the weighted sum of a content loss and an adversarial loss component as:

$$I^{SR} = l_x^{SR} + 10^{-3} l_G^{SR} \quad (3)$$

Where l^{SR} is content loss which can be measured by using MSE loss, l_G^{SR} is the adversarial loss.

The pixel-wise MSE loss can be calculated as:

$$l_{MSE}^{SR} = \frac{1}{r \times W \times H} \left\| I_{x,y}^{HR} - G_{\theta G}(I^{LR}) \right\|^2 \quad (4)$$

In additional to the content loss, we also added the generative component for GAN loss. The generative loss I_G^{SR} is denfined based of the probabilities of discriminator $D_{\theta D}(G_{\theta G}(I^{LR}))$ over all training samples as:

$$l_G^{SR} = \sum_{n=1}^{N} -log D_{\theta D}(G_{\theta G}(I^{LR})) \quad (5)$$

Where $D_{\theta D}(G_{\theta G}(I^{LR}))$ is the probability, so that the reconstructed SR image $G_{\theta G}(I^{LR})$ is a real HR image. For better gradient behavior we minimized $\sum_{n=1}^{N} -log D_{\theta D}(G_{\theta G}(I^{LR}))$ instead of $\log[1 - D_{\theta D}(G_{\theta G}(I^{LR}))]$ (Goodfellow I., 2014).

2.3 Image Quality Assessment

Image quality refers to visually significant attributes of images and focuses on the perceptual assessments of human viewers (Wang, Z. et al, 2020). In general, image quality assessment include subjective methods based on human observer's perceptual evaluation and objective methods based on computational models. Although objective methods can give us quantitative and comparable result, sometime it often unable to capture the human visual perception very accurately. Thus, we chose several commonly used image quality assessment methods, which cover both subjective methods and objective methods.

2.3.1 Peak Signal-to-Noise ratio

Peak signal-to-noise ratio (PSNR) is a metric widely used to assess reconstruction quality from lossy transformation. For remote sensing image super resolution, PSNR is defined via the maximum possible pixel value and the mean squared error (MSE) between images. Give a ground truth image I^{HR} and the reconstruction image I^{SR}, the MSE and PSNR (in dB) can be described as follows:

$$\mathrm{MSE} = \frac{1}{N}\sum_{i=1}^{N}(I^{HR}(i) - I^{SR}(i))^2 \quad (6)$$

$$\mathrm{PSNR} = 10 \cdot \log_{10}\left(\frac{L^2}{MSE}\right) \quad (7)$$

Where L is the maximum possible pixel value of the image (when pixel is represented using 8 bits per smaple, this is 255).

2.3.2 Structural similarity

The structural similarity index (SSIM) is proposed for measuring the structural similarity between images, based on luminance, contrast, and structure (Wang, Z. et al 2004). For an image I with N pixels, the luminance and contrast are estimated as the mean and the standard deviation of the image intensity, respectively can be described as follow:

$$\mu_I = \frac{1}{N}\sum_{i=1}^{N} I(i) \quad (8)$$

$$\sigma_I = \left(\frac{1}{N-1}\sum_{i=1}^{N}(I(i) - \mu_I)^2\right)^{1/2} \quad (9)$$

Where I(i) represents the intensity of i-th pixel of image I. And the comparison functions on luminance and contrast, denoted as $\mathrm{C}_l(I^{HR}, I^{SR})$ and $\mathrm{C}_c(I^{HR}, I^{SR})$ respectively, can be calculated as :

$$\mathrm{C}_l(I^{HR}, I^{SR}) = \frac{2\mu_{I^{HR}}\mu_{I^{SR}} + C_1}{\mu_{I^{HR}}{}^2 + \mu_{I^{SR}}{}^2 + C_1} \quad (10)$$

$$\mathrm{C}_c(I^{HR}, I^{SR}) = \frac{2\mu_{I^{HR}}\mu_{I^{SR}} + C_c}{\mu_{I^{HR}}{}^2 + \mu_{I^{SR}}{}^2 + C_c} \quad (11)$$

Where $\mathrm{C}_1 = (k_1, L)^2$ and $\mathrm{C}_2 = (k_2, L)^2$ are constants for avoiding instability,$k_1 \ll 1$ *and* $k_2 \ll 1$ are small constants, and L is the maximum possible pixel value.

The structural comparison function $\mathrm{C}_s(I^{HR}, I^{SR})$ is defined as:

$$\sigma_{I^{HR}I^{SR}} = \frac{1}{N-1}\sum_{i=1}^{N}(I^{HR}(i) - \mu_{I^{HR}})(I^{SR}(i) - \mu_{I^{HR}})$$

$$\mathrm{C}_s(I^{HR}, I^{SR}) = \frac{\sigma_{I^{HR}I^{SR}} + C_3}{\sigma_{I^{HR}}\sigma_{I^{SR}} + C_3} \quad (12)$$

Where $\sigma_{I^{HR}I^{SR}}$ is the covariance between I^{HR} and I^{SR}, and C_3 is a constant for stability.

Finally, the SSIM can be described by:

$\mathrm{SSIM}(\mathrm{I}^{HR}, \mathrm{I}^{SR}) = [\mathrm{C}_l(\mathrm{I}^{HR}, \mathrm{I}^{SR})]^{\alpha}[\mathrm{C}_c(\mathrm{I}^{HR}, \mathrm{I}^{SR})]^{\beta}[\mathrm{C}_s(\mathrm{I}^{HR}, \mathrm{I}^{SR})]^{\gamma}$

Where α, β, γ are control parameters for adjusting the relative importance. In practice, researcher often set $\alpha = \beta = \gamma = 1$ and $\mathrm{C}_3 = \mathrm{C}_2/2$. (Brunet, D. et al. 2012)

2.3.3 Mean Opinion Scores

Mean opinion scores (MOS) testing is a commonly used subjective IQA (Image Quality Assessment) method. It is the arithmetic mean over all individual values on a predefine scale

that a subject assigns to the opinion of the performance of a system quality (Q. Huynh-Thu, 2011). Specifically, we asked 15 raters to assign an integral score from 1 (bad quality) to 5 (excellent quality) for the performance of tested images. And the final MOS is calculated as the arithmetic mean over single ratings performed by human subjects for given stimulus, which can be described by:

$$MOS = \frac{\sum_{n=1}^{n} R_n}{N} \quad (13)$$

Where R are the individual ratings for a given stimulus by N subjects.

Table 2. Comparison of NN, bicubic, SRCNN, our method and original HR on GF2 satellite image.

Test data regions	Methods	Nearest	Bicubic	SRCNN	Ours	Original
Resident	PSNR	25.98	27.13	29.81	29.93	∞
	SSIM	0.7332	0.8037	0.8409	0.8733	1
	MOS	1.12	1.88	3.37	4.12	4.33
River	PSNR	24.77	27.48	29.30	29.17	∞
	SSIM	0.7636	0.8456	0.8542	0.8957	1
	MOS	1.12	3.38	4.32	4.32	4.76
Farmland	PSNR	21.39	22.57	24.31	24.65	∞
	SSIM	0.6931	0.7358	0.7343	0.7356	1
	MOS	1.33	1.83	3.69	4.19	4.41

Figure2. Comparison of using bicubic interpolation, SRCNN, our SR-GAN method and original HR image

3. EXPERIMENTS

3.1 Experimental Data

The learning-based SR methods learn the missing high-frequency information of LR image from provided in the training HR data. Thus, beside the mount of sample data, the performance of the SR reconstruction method is also related to the similarity of the test image to the training image. To obtain a more targeted training performance, the type of earth surface should also be taken into consideration.

In this work, our goal is to reconstruct satellite image for the most common ground objects. Therefore, we use GF-2 satellite images (GSD 1m) as an example for rural area. It includes common ground objects such as river, resident, farmland and so on. These HR images of GF-2 satellite were down sampled to LR images and used as training samples. And their original HR images are regarded as their ground truth labels. We split this image into patches with size of 250×250 pixels and1000 patches of them were randomly selected. They were further grouped into 800, 100 and 100 patches, which were used as training dataset, validation dataset, and test samples respectively.

All experiments were performed with a scale factor of 4× between low- and high-resolution images. Super resolved images for the reference methods, including nearest, bicubic, SRCNN (Kim et al., 2016) and For fair comparison, all results PSNR and SSIM measures were calculated.

3.2 Training detail and parameters

As a supervised method, training is a necessary process for the obtainment of a robust generative adversarial network for satellite image super resolution. Because of our network architecture, training processes can be divided into two parts.

For the generator network, ResNet with extremely deep layers (more than 1000 layers) have been proved which can extract and represent more features and semantic information beyond the other deep learning network architecture recently (He, Kaiming, et al., 2016). However, because of the huge numbers of weights parameters in ResNet, it needs tremendous amount of sample data and computational resource. Thus, we used a pre-trained ResNet model that was previously trained on ImageNet datasets (Deng, Jia, et al, 2009) a large dataset consisting of 1.4M images and 1000 classes. Then, we retrained the weights of the top layers of the ResNet pre-trained model alongside the training of the generative high resolution images we need. The training process will force the weights to be tuned from generic feature maps to features associated specifically with the satellites dataset.

For the discriminator network, the training process is a min-max algorithm between the generator G and the discriminator D. The generator G takes generative high resolution image as inputs, to make the generative high resolution images more realistic, the generator G seeks to minimize loss functions. Additional discriminator network D tries its best to distinguish the differences between generative high resolution image and original high resolution by maximizing loss function. The Training process employs the Adam optimizer for computing the minimization and maximization. The Adam optimizer makes the training of our model converge fast with economical computing resources. In addition, the Adam optimizer employs an independent adaptive learning rate strategy, which enables the computation for large-scale parameter optimization even more efficient. The training procedures can be found in table 1.

Table 1. Training procedure for our satellite image super resolution

INPUT: low resolution image , original high resolution image
For the number of training iterations DO
UPDATE G:
Take mini-batch examples from low resolution images to do super resolution
Update parameters of G via minimizing the sum of (4) by using the Adam optimizer
UPDATE D:
Take generative high resolution image and original high resolution image as input.
Update parameters of D via maximizing (5) by using the Adam optimizer.
End
OUTPUT: generator network G

When the discriminator D cannot distinguish the differences between generative high resolution image and original resolution image, the generator network G wins the min-max adversarial game, which means the well trained generator network G is good enough to reconstruct realistic looking super resolution image from low resolution image.

We trained all networks on a NVIDIA Titan X (Pascal) GPU using image batch size is 16, these images are distinct from the testing images. For optimization we use Adam (Kingma et al., 2015) with β_1 = 0.9. The SRCNN networks were trained with learning rate of 10^{-3} and 10^5 update iterations. The EDSR network was trained for 10^5 update iterations and evaluation every 5000 iterations. Our GAN network used the MSE-based SRCNN network as initialization for generator during training to avoid undesired local optima. The GAN model variants were trained with 105 iterations. The implementation of entire models is based on Tensorflow 2.0.

3.3 Complementarity Analysis of SR-GAN network

For the sake of comprehensiveness, we conducted subjective qualitative and objective quantitative analyses on reconstruct SR images which be generated by using different SR method. We compared the performance of our SR-GAN and nearest, bicubic interpolation, SRCNN (Dong, C. et al., 2015).

From an objective perspective, PSNR and structural similarity (SSIM) index are used to evaluate the reconstruction result. Quantitative results are summarized in Table 1; our method yields the highest scores in evaluation matrices among all SR methods in both PSNR and SSIM. It is worth to point out that these is not significant an advantage over the SRCNN in PSNR matrices. This is caused by competition between MSE based content loss in SRCNN and the adversarial loss in our method.

To evaluate the subjective performance of difference SR methods, we chose five ground objects with representative scale, the visual examples provided in Figure 2. Our method produced sharper edges in building roof and street corner, better detail texture in farmland. Especially, as displayed in Figure 3, for vehicle and building with more complex textures in image, our method kept more detail texture from original HR image. In general, high-spatial-resolution image reconstructed by our method create more high-frequency information.

* Corresponding author.

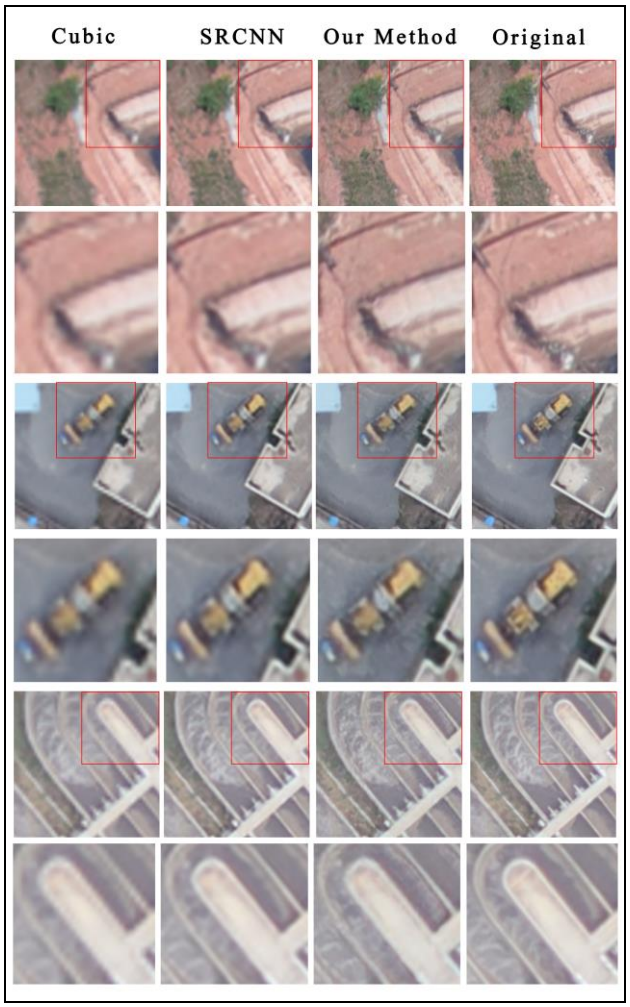

Figure 3. The local image details from SR reconstruction results and corresponding reference HR image

4. CONCLUSION AND FUTURE WORKS

This paper presents a generative adversarial network for single satellite image Super Resolution reconstruction. It built on a trained deep residual network to generate preliminary SR images, combined with a discriminative network learns to differentiate preliminary SR images and High resolution samples. The experiments results show that our method can use existing model parameters to refine SR image performance. Compare to previously published learning based SR methods, in which the obtainment of HR image mainly relies on mount of image sample data. And the characteristics of trained satellite image sample is determined the performance of SR reconstruction.

From a perspective of application, the reconstruction result of GF2 images proven our SR-GAN method is useful in the obtainment of high-resolution images. In future, we will establish a more general SR-GAN model to improve the performance of Super resolution with different satellite images.

ACKNOWLEDGEMENTS

This work was supported by the Special Project of Science and Technology Basic Resources Survey, China Ministry of Science and Technology, under Grant 2019FY202502.

REFERENCES

Benediktsson, Jón Atli, Jocelyn Chanussot, and Wooil M. Moon., Advances in very-high-resolution remote sensing. Proc. IEEE 101.3 (2013): 566-569.

Brunet, Dominique, Edward R. Vrscay, and Zhou Wang. "On the mathematical properties of the structural similarity index." IEEE Transactions on Image Processing 21.4 (2011): 1488-1499.

Deng, Jia, et al. "Imagenet: A large-scale hierarchical image database." 2009 IEEE conference on computer vision and pattern recognition. Ieee, 2009.

Dong, C., Loy, C. C., He, K., & Tang, X. (2015). Image super-resolution using deep convolutional networks. IEEE transactions on pattern analysis and machine intelligence, 38(2), 295-307.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-adversarial nets. In Advances in neural information processing systems (pp. 2672-2680).

H. Ahn, B. Chung and C. Yim, "Super-Resolution Convolutional Neural Networks Using Modified and Bilateral ReLU," 2019 International Conference on Electronics, Information, and Communication (ICEIC), 2019, pp. 1-4, doi: 10.23919/ELINFOCOM.2019.8706394.

He, Kaiming, et al. "Deep residual learning for image recognition." Proceedings of the IEEE conference on computer vision and pattern recognition. 2016.

Kingma, Diederik P., and Jimmy Ba. "Adam: A method for stochastic optimization." In International Conference on Learning Representations (ICLR), 2015. 6

Kim, J., Kwon Lee, J., & Mu Lee, K. (2016). Accurate image super-resolution using very deep convolutional networks. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 1646-1654).

Ledig, C., Theis, L., Huszár, F., Caballero, J., Cunningham, A., Acosta, A., ... & Shi, W. (2017). Photo-realistic single image super-resolution using a generative adversarial network. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 4681-4690).

Q. Huynh-Thu, M. Garcia, F. Speranza, P. Corriveau and A. Raake, "Study of Rating Scales for Subjective Quality Assessment of High-Definition Video," in IEEE Transactions on Broadcasting, vol. 57, no. 1, pp. 1-14, March 2011, doi: 10.1109/TBC.2010.2086750.

Lim, K. H., & Kwoh, L. K. (2009). Super-resolution for SPOT5-Beyond supermode. In 30th Asian Conference on Remote Sensing (pp. 378-382).

Lu, T., Wang, J., Zhang, Y., Wang, Z., & Jiang, J. (2019). Satellite Image Super-Resolution via Multi-Scale Residual Deep Neural Network. Remote Sensing, 11(13), 1588.

Luo, Y., Zhou, L., Wang, S., & Wang, Z. (2017). Video satellite imagery super resolution via convolutional neural networks. IEEE Geoscience and Remote Sensing Letters, 14(12), 2398-2402.

Lim, B., Son, S., Kim, H., Nah, S., & Mu Lee, K. (2017). Enhanced deep residual networks for single image super-resolution. In Proceedings of the IEEE conference on computer vision and pattern recognition workshops (pp. 136-144).

Nasrollahi, K., & Moeslund, T. B. (2014). Super-resolution: a comprehensive survey. Machine vision and applications, 25(6), 1423-1468.

Radford, A., Metz, L., & Chintala, S. (2015). Unsupervised representation learning with deep convolutional generative adversarial networks. arXiv preprint arXiv:1511.06434.

Schulter, S., Leistner, C., & Bischof, H. (2015). Fast and accurate image upscaling with super-resolution forests. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 3791-3799).

Wang, Z., Bovik, A. C., Sheikh, H. R., & Simoncelli, E. P. (2004). Image quality assessment: from error visibility to structural similarity. IEEE transactions on image processing, 13(4), 600-612.

Wang, Z., Chen, J., & Hoi, S. C. (2020). Deep learning for image super-resolution: A survey. IEEE Transactions on Pattern Analysis and Machine Intelligence.

Yang, J., Wright, J., Huang, T. S., & Ma, Y. (2010). Image super-resolution via sparse representation. IEEE transactions on image processing, 19(11), 2861-2873.

Yu, Y., Gong, Z., Zhong, P., & Shan, J. (2017, September). Unsupervised representation learning with deep convolutional neural network for remote sensing images. In International Conference on Image and Graphics (pp. 97-108). Springer, Cham.