Enhancing the Resolution of Satellite Imagery Using a Generative Model

Mahee Tayba * and Pablo Rivas †, *Senior, IEEE*
School of Engineering and Computer Science
Department of Computer Science, Baylor University
Email: {*Mahee_Tayba1,†Pablo_Rivas}@Baylor.edu

***Abstract*—Recent breakthroughs in deep learning algorithms introduced the image super-resolution technique that maps the low-resolution image to generate a high-resolution image. These techniques increase various surveillance applications by providing finer spatial details than data from original sensors. Satellite images obtained from Moderate Resolution Imaging Spectroradiometer (MODIS) observation offer essential information about the earth's landscape, ocean, and ecosystem, contributing to monitoring various applications in the scientific field. The spatial resolution of satellite images has a significant impact on image accuracy. This paper focuses on improving image resolution by training a convolutional neural network to produce super-resolution images from low-resolution images. We present an implementation of Super Resolution Generative Adversarial Network (SRGAN), a GAN-based approach that uses a perceptual loss function that includes an adversarial loss and a content loss. Using a discriminator network that is designed for discerning between super-resolved images and original photo-realistic images, the adversarial loss drives the solution of this architecture to natural images. Moreover, the content loss is driven by perceptual similarity rather than pixel space similarity. We used this architecture to satellite images collected from NASA MODIS devices and found satisfactory results. Our key finding is that our system's result can now be used to improve a variety of low-resolution images.**

***Index Terms*—image resolution, GAN, satellite image, deep neural network, super resolution**

I. Introduction

In the past few years, satellite imagery processing has received significant attention in the scientific community's significance in a wide range of application circumstances. A variety of real-world applications, including medical images for clinical activities, aerial image interpretation, geographic image analysis, video surveillance, astronomical data analysis, require high-resolution images. As high resolution (HR) images provide critical information, they are also highly desirable in satellite images. Satellite imaging with high resolution can capture fine details on the surface. Thousands of artificial satellites are currently orbiting the earth, each taking photographs of space for a different purpose. These satellites keep an eye on clouds, wildfires, volcanoes, oceans, land, and ice. Land analysis, weather and disaster prediction, air quality measurement, topographic analysis, geosciences, and other applications depend heavily on satellite images, and those images have numerous applications in scientific research studies.

For capturing detailed information, improving the resolution and accuracy of an image is essential. The task of reconstructing high-resolution images from one or more low-resolution (LR) images of the same scene is known as super-resolution (SR). To put it another way, LR stands for single image input, HR stands for the actual image, and SR stands for the predicted high-resolution image. The LR images are usually down-sampled HR photos with some noise applied when using machine learning algorithms. The SR can be categorized into single image super-resolution (SISR) and multi-image super-resolution (MISR) based on the number of input LR images. Satellite imaging is an excellent case where SISR may be helpful.

In recent decades, enhacing image quality has been a popular research area in the study of image processing and computer vision. Deep Convolutional Neural Networks (CNNs) [1] are often used to solve the challenge of image super-resolution since they exhibit remarkable accuracy improvements. Reconstruction based techniques for superresolving satellite images have been developed in the recent years [2], [3]. The residual learning technique was incorporated into computer vision tasks to develop a deeper CNN demonstrating outstanding achievement [4]. GAN [5] has also been utilized to image SR in computer vision. GAN is a deep convolutional network that acts as a strong tool for building natural-looking images with high image quality, which pushes reconstructions to travel toward search space regions with a high likelihood of having photo-realistic images, bringing them closer to the natural image spectrum. The goal of this paper is to enhance the resolution of satellite images by using the SRGAN proposed by Ledig *et al.* in [6].

The main contribution of this paper is implementing a resolution enhancement technique using an adversarial learning process on satellite images to increase the image resolution. The performance of the method is then measured using the PSNR (power signal-to-noise ratio) method. The result shows that the image resolution has been improved without losing quality. The results and discussions section contains both visual and quantitative information.

Section II presents a brief summary of related work in image super-resolution, specially applied to satellite data and the architecture of GAN and VGG19. Section III discusses the methods, whereas Section IV discusses the data. The experimental setup and model are described in Section V. The results are discussed in Section VI, and the paper's primary conclusions are drawn in Section VII.

II. Related Work

The most deciding feature of an image has been defined as "resolution" in many video processing applications [8] and satellite image resolution enhancement [9]. Various image resolution enhancement techniques came into the picture as high-resolution images are expensive to capture and hard to obtain due to certain limitations in terms of using image sensors and physical devices [10]. For super resolution images, both interpolation methods [11] and learning based algorithms [12] are used as reconstruction method. [13] showed a Wavelet-based resolution enhancement technique to enhance image resolution using 250m resolution MODIS data. Super resolution (SR) algorithms based on convolutional neural networks(CNN) [14] have achieved great accomplishment in recent years. SRCNN [15] learns an end-to-end mapping connection from LR to HR using a three-layer network hierarchy. More sophisticated networks are proposed to increase accuracy even more. [16] expanded the layer count to 20 and used tiny filters, a fast learning rate, and gradient clipping that could be adjusted. [17] proposed an enhanced classification of satellite images that uses CNNs. [6] uses three loss functions (perceptual loss, adversarial loss, and content loss) to achieve significant results while improving image quality. They retrieved photo-realistic textures from heavily downsampled images. [18] achieved good results over ZSSR (zero-shot super-resolution) and bicubic interpolation. To identify and analyze objects from satellite images [19], Generative Adversarial Network(GAN) has also obtained great accuracy.

A. GAN by Goodfellow et al.

One of the most basic neural-based models that implement adversarial learning is a GAN [5]. It trains two neural network at the same time. A generative model G, that captures the data distribution, and a discriminative model, D that tries to determine whether the sample coming from training data or G. The generator creates random noise from a random distribution, and an adversary is created to produce fake inputs. Then the discriminative model tries to learn to distinguish if the sample coming from training data is true or fake input. To properly recognize true input, the model learns from its mistake by adjusting itself. In order to make the model fail, the adversary would need to keep generating false inputs.

B. VGG19 by Simonyan et al.

VGG19 [7] is a 19-layer deep convolutional neural network, which is a pre-trained model which has already been trained on a large dataset ImageNet [20] which has more than a million images. This network was evaluated by pushing a 3×3 layer to 19 layers which is showed in Fig. 1. The authors started by adding more layers to VGGNet to get less error rate, but after VGG16 the error rate did not improve, so they stopped at adding 19 layers. This network has learned a diverse number of rich feature representations for a variety of images, which can classify a number of different object categories.

C. SRGAN by Ledig et al.

The functionality of the generator and discriminator of the GAN is as follows:

The generator network takes input image and after a series of convolutional layers and upsampling layers, it generates a super-resolution image of shape $256 \times 256 \times 3$. The discriminator takes a high-resolution image and tries to identify whether the image is generated or real(from data sample).

1) The Generator Network architecture: It contains the following blocks which is showed in Fig. 2.

The pre-residual block It contains a single 2D convolutional layer with an input shape of (64,64,3) and PReLU as an activation function. The output shape has a feature size of 64.

The residual block The generator contains 16 residual blocks which use Adam optimizer. The residual block contains two 2D convolutional layers which are followed by a batch normalization layer. There is also an extra layer that calculates the sum of the input to the residual block and the output of the batch normalization layer. *The post residual block* It contains a single 2D convolution layer and PReLU as the activation function which is followed by a batch normalization layer. *The upsampling block* There are two upsampling blocks in the generator. Each of them contains one convo 2D layer and one upsampling 2D layer that uses PRelu as an activation function. The resolution of the input image is increased and the final output shape is (256, 256, 256). *The final convolution layer* To generate the shape of the input image back which is (256, 256, 3), the last convolution 2D layer is applied.

2) The Discriminator Network architecture: Discriminator determines the validity of generated high-resolution images. This deep convolutional network contains 8 convolution layers, each with an increasing number of 3×3 filter kernels, rising by a factor of two from 64 to 512 kernels [6]. Each time the features are doubled, stridden convolutions are used to reduce the image resolution. Each convolution block is followed by a batch normalization layer. The two dense layers act as classification blocks and the last layer predicts the chance of an image being fake or real. The architecture is showed in Fig. 3.

III. Methodology

For super image resolution, we will use SRGAN, which is capable of inferring photo-realistic natural images [6]. SRGAN exploits a multi-task loss to refine the result of a GAN-based architecture that uses the SRResnet network architecture as the basic building block.

We obtained low-resolution images by resizing the high-resolution input images. It was reduced by 4 times. So, the input 256×256 images become 64×64 low-resolution images. Now the generator produces fake high resolution (256×256) images. At this point, the discriminator tries to identify the discriminator loss between input high-resolution and fake high-resolution images. Then, we extracted the features of high-resolution images by using a pre-trained VGG19 network and used it to train the generator against low-resolution images along with high-resolution images.

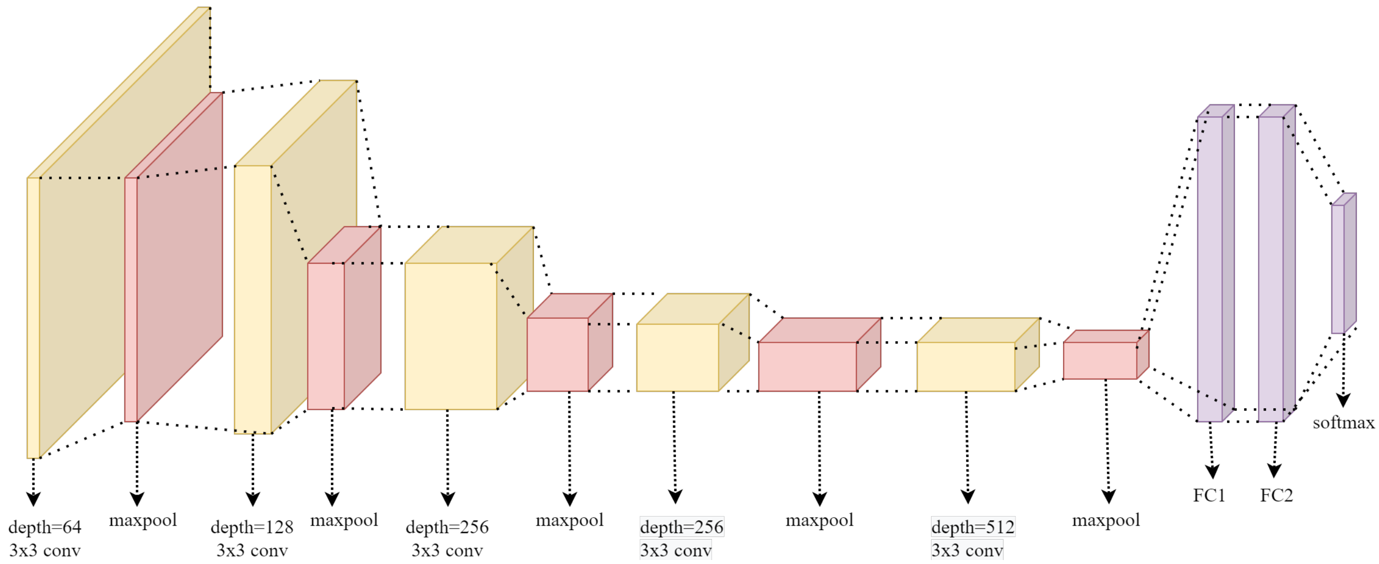

Fig. 1. The 19 layers of VGG19 network used for feature extraction. This architecture was initially proposed by Simonyan *et al.* in [7]

.

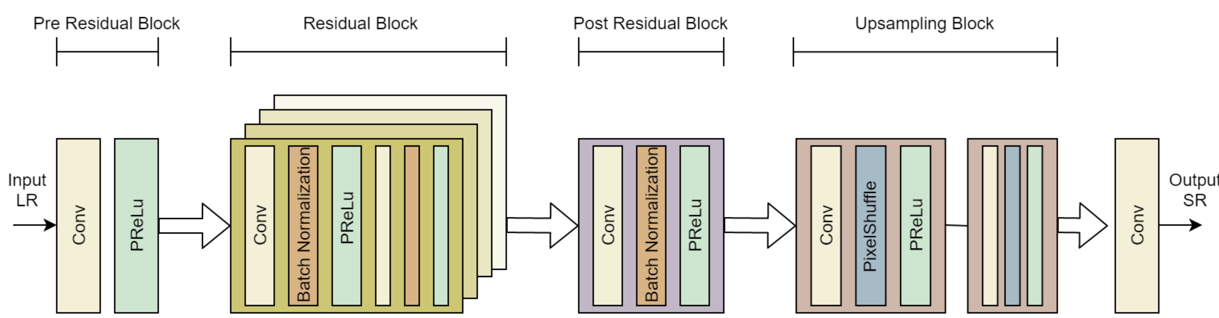

Fig. 2. Architecture for Generator network.

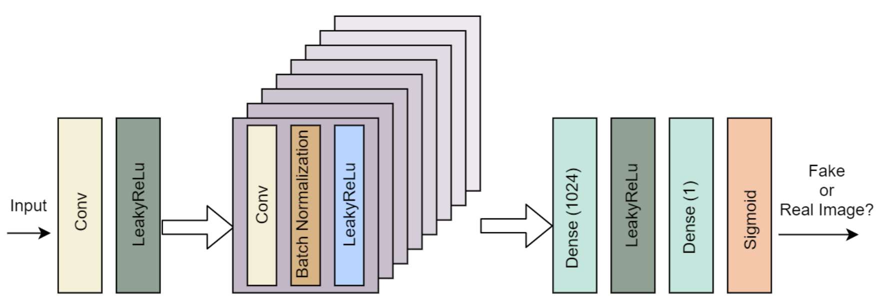

Fig. 3. Architecture for Discriminator network.

A. Data

The data used in this project was collected from the Moderate Resolution Imaging Spectroradiometer (MODIS) instrument provided by National Aeronautics and Space Administration (NASA). Both NASA Terra and Aqua spacecraft have the MODIS instrument mounted on them. It has a viewing width of 2,330 km and takes one to two days to view the entire surface of the earth by collecting information by a wide-angle lens and in 36 spectral bands. MODIS data is being used by researchers to improve our understanding of global dynamics and climate change. Each pixel in the most accurate NASA images measures 10 meters per pixel, our MODIS data has a spatial resolution of 250m.

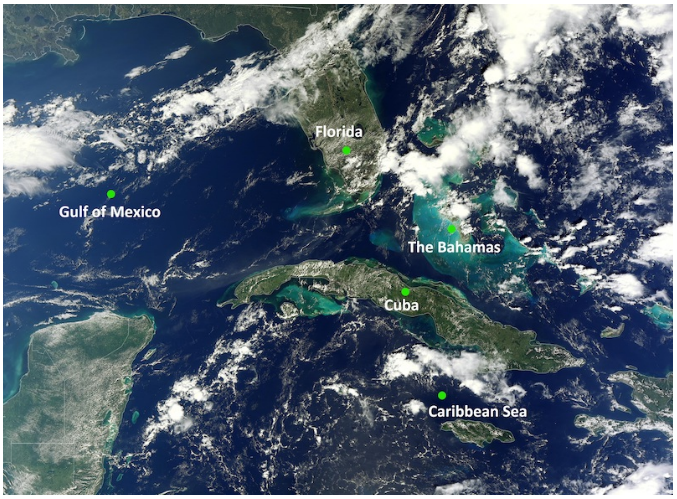

Fig. 4. Input data of southeast USA, Ocotber 19, 2012, 16:05 UTC. Satellite: Terra. Instrument: MODIS. Spatial resolution: 250m

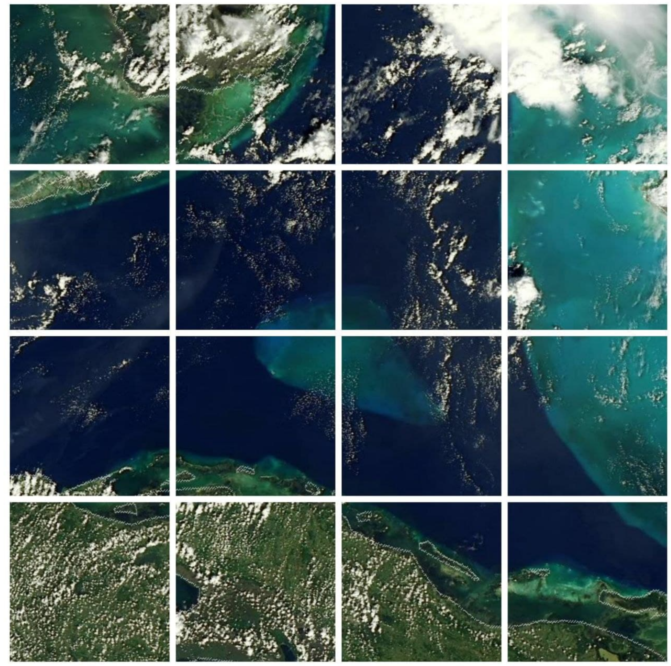

Fig. 5. A sample data of satellite image that is cropped from center.

We used celebA dataset [21] for initial training, which is widely used in detecting, identifying, and performing face recognition systems in many computer applications. This dataset has more than two hundred thousand images each with 40 attribute annotations of the faces of different celebrities.

B. Experiments

At first, we implemented the methods from [6] by using CelebFaces Attributes (CelebA) dataset,[1] then we replicated this work in Google Colab by upgrading it in Tensorflow v2 and then used that model to our satellite dataset. We trained the model for first 5000 epoch on celebA dataset. The training took more than two days on a single GPU computer. We have seen the phenomenon that, if we feed something different than face to the system, the resulting image tries to generate a face. So we trained the model on our satellite image for the rest 20000 epochs and the system predicted the high resolution images correctly.

We conducted data augmentation to increase the size of our dataset. One sample data has been showed in Fig. 4. As the center has better resolution and clarity, we chose 1024×1024 from the center and took 16 images each sized 256×256, which led us to 1248 images which were initially 78. Some of the images showed black borders, so after removing 8 images, and finally worked with 1240 images. We used the bi-cubic interpolation to resize the images. After cropping and resizing, the images looked like Fig. 5.

IV. Result and Analysis

The power signal-to-noise ratio (PSNR) measures the ratio of highest pixel intensity to distortion power, and it is generated from mean squared error (MSE). It can be defined as follows:

$$PSNR = 20 \cdot \log_{10}(MAX_I) - 10 \cdot \log_{10}(MSE), \quad (1)$$

[1]https://mmlab.ie.cuhk.edu.hk/projects/CelebA.html

where MAX_I denotes the maximum intensity in the image. The PSNR is one of the preferred measures of distortion in reconstruction of images [22]. Fig. 6 shows an original cropped satellite image (right), which gets downscaled and is reconstructed back to its original high-resolution using the SRGAN (middle) and bicubic interpolation (left). The figure includes the PSNR to show that the model learned produces a higher PSNR than bicubic interpolation [23]. Fig. 6 also shows two different samples; on the top, it shows an image with land and ocean, and at the bottom, it shows an image with clouds and vegetation. From the image, we can see that the resolution has been improved without losing quality.

The authors in [6], showed that as the number of training iterations increases, the SRGAN generator network improves its reconstruction performance. For this reason, we continued to train our models for extended periods, producing the results depicted in Fig. 7. The figure shows the difference between the actual and the generated high-resolution image in different image regions. For comparison, we show the original and the reconstructed version from a low-quality input, and our results indicate that the model performs well with very minimal visual differences that are only perceived on zoomed-in versions.

The results obtained visually and numerically by showing picture differences and PSNR amounts suggest that SRGANs are a viable resource to increase the resolution of satellite imagery. The experiments conducted here included downscaling of up to a factor of four. In addition, the results show that the models can execute proper reconstructions from low-resolution imagery.

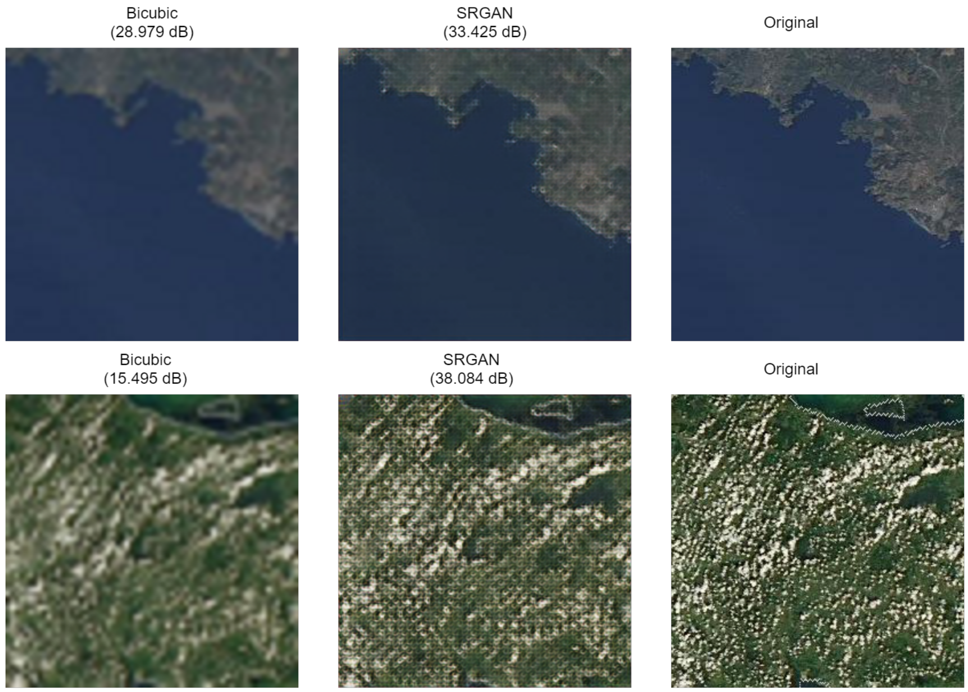

Fig. 6. PSNR evaluation of SRGAN for MODIS dataset with 250m spatial resolution. Right: original images that are down-scaled by a factor of four. Middle: SRGAN reconstruction form down-scaled image. Left: Bicubic interpolation reconstruction from down-scaled image. Top and Bottom are different satellite readings.

Fig. 7. Results after SRGAN reconstruction and corresponding reference HR image. Differences are barely noticeable in zoomed-in versions on the right.

V. Conclusion

We trained a GAN to perform super-resolution image reconstructions on satellite images. The SRGAN augments the traditional reconstruction loss function with an adversarial loss. The experiment results show that applying super-resolution GAN to improve image resolution leads to photorealistic and clear images, an essential contribution in satellite image analysis and processing. Furthermore, the model outperforms the classic bicubic method when evaluated visually and quantitatively with a PSNR measure.

We believe that the main contribution of this paper is to show an interesting application for enhancing resolution of satellite imagery, based on the SRGAN initially proposed by Ledig *et al.* in [6]. Increasing resolution based on learned characteristics of satellite imagery can lead to automatic image enhancement based on evidence rather than numerical guess. Scientific analysis of satellite data can be potentially improved to include more detailed regional analysis on smaller areas.

Future work includes using more significant amounts of satellite images with varied regions all over the world. Our initial experiments were conducted on a limited set to retain control of our experiments. Furthermore, SRGANs will be trained on other available satellite datasets that could be helpful for future scientific research. The current experiment was limited to using MODIS data.

Acknowledgment

The work presented here was supported in part by the Baylor AI lab in Baylor University's Department of Computer Science.

References

[1] B. Lim, S. Son, H. Kim, S. Nah, and K. Mu Lee, "Enhanced deep residual networks for single image super-resolution," in *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, 2017, pp. 136–144.

[2] F. Li, X. Jia, and D. Fraser, "Universal hmt based super resolution for remote sensing images," in *2008 15th IEEE International Conference on Image Processing*. IEEE, 2008, pp. 333–336.

[3] M. T. Merino and J. Nunez, "Super-resolution of remotely sensed images with variable-pixel linear reconstruction," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 45, no. 5, pp. 1446–1457, 2007.

[4] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2016.

[5] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial networks," *arXiv preprint arXiv:1406.2661*, 2014.

[6] C. Ledig, L. Theis, F. Huszár, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. 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.

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

[8] T. Celik, C. Direkoglu, H. Ozkaramanli, H. Demirel, and M. Uyguroglu, "Region-based super-resolution aided facial feature extraction from low-resolution sequences," in *Proceedings.(ICASSP'05). IEEE International Conference on Acoustics, Speech, and Signal Processing, 2005.*, vol. 2. IEEE, 2005, pp. ii–789.

[9] H. Demirel and G. Anbarjafari, "Satellite image resolution enhancement using complex wavelet transform," *IEEE geoscience and remote sensing letters*, vol. 7, no. 1, pp. 123–126, 2009.

[10] S. C. Park, M. K. Park, and M. G. Kang, "Super-resolution image reconstruction: a technical overview," *IEEE signal processing magazine*, vol. 20, no. 3, pp. 21–36, 2003.

[11] J. Yang, J. Wright, T. S. Huang, and Y. Ma, "Image super-resolution via sparse representation," *IEEE Transactions on Image Processing*, vol. 19, no. 11, pp. 2861–2873, 2010.

[12] S. Schulter, C. Leistner, and H. Bischof, "Fast and accurate image upscaling with super-resolution forests," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 3791–3799.

[13] A. Garg, S. V. Naidu, T. Ahmed, H. Yahia, and D. Singh, "Wavelet based resolution enhancement for low resolution satellite images," in *2014 9th International Conference on Industrial and Information Systems (ICIIS)*. IEEE, 2014, pp. 1–5.

[14] C. Dong, C. C. Loy, K. He, and X. Tang, "Learning a deep convolutional network for image super-resolution," in *European conference on computer vision*. Springer, 2014, pp. 184–199.

[15] ——, "Image super-resolution using deep convolutional networks," *IEEE transactions on pattern analysis and machine intelligence*, vol. 38, no. 2, pp. 295–307, 2015.

[16] J. Kim, J. K. Lee, and K. M. Lee, "Accurate image super-resolution using very deep convolutional networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 1646–1654.

[17] N. Laban, B. Abdellatif, H. M. Ebied, H. A. Shedeed, and M. F. Tolba, "Performance enhancement of satellite image classification using a convolutional neural network," in *International Conference on Advanced Intelligent Systems and Informatics*. Springer, 2017, pp. 673–682.

[18] L. Zhao, H. Yu, and Y. Wang, "Mslp: Deep superresolution for meteorological satellite image," *Complexity*, vol. 2021, 2021.

[19] P. Pushkar, L. Aggarwal, M. Saad, A. Maheshwari, H. Awasthi, and P. Nagrath, "Object identification in satellite imagery and enhancement using generative adversarial networks," in *International Conference on Innovative Computing and Communications*. Springer, 2021, pp. 643–657.

[20] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in *2009 IEEE Conference on Computer Vision and Pattern Recognition*, 2009, pp. 248–255.

[21] Z. Liu, P. Luo, X. Wang, and X. Tang, "Large-scale celebfaces attributes (celeba) dataset," *Retrieved August*, vol. 15, no. 2018, p. 11, 2018.

[22] A. Hore and D. Ziou, "Image quality metrics: Psnr vs. ssim," in *2010 20th international conference on pattern recognition*. IEEE, 2010, pp. 2366–2369.

[23] D. Han, "Comparison of commonly used image interpolation methods," in *Proceedings of the 2nd International Conference on Computer Science and Electronics Engineering (ICCSEE 2013)*, vol. 10, 2013.