Proceeding Paper

Generating Super Spatial Resolution Products Form Sentinel-2 Satellite Images †

Mohammad Reza Zargar and Mahdi Hasanlou *

School of Surveying and Geospatial Engineering, College of Engineering, University of Tehran, Tehran 14174-66191, Iran; hasanlou@ut.ac.ir

* Correspondence: hasanlou@ut.ac.ir; Tel.: +98-21-6111-4525 (M.H.)

† Presented at the 5th International Electronic Conference on Remote Sensing, 7–21 November 2023; Available online: https://ecrs2023.sciforum.net/

Abstract: Access to high spatial resolution satellite images enables more accurate and detailed analysis of these images. Furthermore, it facilitates easier decision-making on a wide range of issues. Nevertheless, there are commercial satellites such as Worldview that have provided a spatial resolution of fewer than 2.0 meters, but using them for large areas or multi-temporal analysis of an area brings huge costs. So, to tackle these limitations and access free satellite images with higher spatial resolution, there are challenges that are known as single-image super-resolution (SISR). The Sentinel-2 satellites were launched by the European Space Agency (ESA) to monitor the Earth which has enabled access to free multi-spectral images, five-day time coverage, and global spatial coverage has been among the achievements of this launch. Also, it led to the creation of a new flow in the field of space businesses. These satellites have provided bands with various spatial resolutions which the Red, Green, Blue, and NIR bands have the highest spatial resolution by 10 m. In this study, therefore, to recover high-frequency details, increase the spatial resolution, and cut down costs Sentinel-2 images have been considered. Additionally, a model based on Enhanced Super-Resolution Generative Adversarial Network (ESRGAN) has been introduced to increase the resolution of 10 m RGB bands to 2.5 m. In the proposed model, several spatial features were extracted to prevent pixelation in the super-resolved image and were utilized in the model computations. Also, since there is no way to obtain higher-resolution (HR) images in the conditions of the Sentinel-2 acquisition image, we preferred instead to simulate data, to use a sensor with a higher spatial resolution that is similar in spectral bands to Sentinel-2 as a reference and HR image. Hence, Sentinel-Worldview image pairs were prepared and the network was trained. Finally, the evaluation of the results obtained, showed that while maintaining the visual appearance, it was able to maintain some spectral features of the image as well. The average Peak Signal-to-Noise Ratio (PSNR), Structural Similarity Index (SSIM), and Spectral Angle Mapper (SAM) metrics of the proposed model from the test dataset were 37.23 dB, 0.92, 0.10 radians, respectively.

Keywords: high spatial resolution; super-resolution; ESRGAN; PSNR; SAM; Sentinel-2.

1. Introduction

Satellite remote sensing is used in various practical domains such as mapping, agriculture, environmental protection, land use, urban planning, geology, natural disasters, hydrology, oceanography, meteorology, and many more. In this field, the spatial resolution is a fundamental parameter that is of particular importance for two reasons. The higher the resolution of an image, the more details and information are available for human interpretation and automated machine understanding [1]. However, the goal of providing high spatial resolution devices at the market price is not always the optimal solution [2]. In the past two decades, super-resolution (SR) has been defined as the process

of reconstructing and generating a high-resolution (HR) image from one or more low-resolution (LR) images. The challenge with the SR process is fundamentally that it is an ill-posed problem. This means that a single LR image can be derived from several HR images with minor variations, such as changes in camera angle, color, brightness, and so on [3]. Therefore, various approaches have been developed with different constraints and multiple equations to map an LR image to an HR image.

Today, the Sentinel-2 constellation freely produces multispectral optical data, providing five-day temporal and global spatial coverage. These achievements have led to the creation of a new flow in the field of space businesses [4].

The SR methods are categorized into multi-image or single-image approaches. In the Multi-Image Super-Resolution (MISR) approach, lost information is extracted from multiple LR images captured of a scene to recover the HR image [5]. However, due to the different angles and passes at which satellite images are acquired for a scene and the sub-pixel misalignments present in these images, the correspondence between a pixel in one image and the same pixel in the next image will not be exact. Therefore, in practical applications, obtaining a sufficient number of images for a scene is challenging [1]. In contrast, in the Single-Image Super-Resolution (SISR) approach, the output HR image is obtained using only one LR image. Therefore, SR can be applied without the need for satellite constellation, which eventually results in substantial cost savings and provides a good opportunity for small platforms, with low resolution and cheap instruments [6].

With the advent of efficient Deep Learning (DL) methods, the SISR approach gradually became dominant among the approaches [7]. Today, one common technique in SR algorithms is utilizing convolutional neural networks (CNNs). The pioneer in CNN-based algorithms was the super-resolution convolutional neural network (SRCNN) [8], and its enhancement has persisted to this day through various methodologies. On the other hand, Generative Adversarial Networks (GANs) [9] have garnered growing interest within the research community. A network based on GANs has been able to generate outputs that are more photo-realistic, albeit with lower quantitative metrics [2].

Satellite imagery, unlike standard imagery, suffers from significant landscape at the pixel level due to its lower resolution. Consequently, satellite imagery, such as Sentinel-2 images, is more prone to pixelation artifacts during the SR process. In this context, we have proposed a model that has been able to mitigate the aforementioned problems using spatial features. In fact, in the proposed network, an attempt has been made to obtain better results in satellite images by modifying the input features of the Enhanced Super-Resolution Generative Adversarial Network (ESRGAN) [10].

The rest of the paper is organized as follows. In Section 2, we describe datasets and the proposed method. Evaluation metrics, comparative solutions, and experimental results are then gathered in Section 3. Finally, Section 4 provides concluding remarks.

2. Materials and Methods

The development of a DL-based method involves at least three key steps, with some iterations among them:

- Preparing an appropriate dataset for training and testing.
- Design and develop a DL model.
- Training and evaluating the model using prepared data.

Following this approach, we will outline the progression of our proposed model in the same order in order to make the material clearer.

2.1. Dataset

In deep learning, data functions as a critical component concerning model quality and efficiency. Deep neural networks necessitate a significant volume of training data to understand intricate patterns. The vast and multifarious nature of data resolves numerous issues associated with model execution and performance decline [11].

We utilized archived WorldView-2/3 images in the European Space Agency (ESA) as ground truth to prepare the LR-HR image pairs. These images were sensed by the satellite from February 2009 to December 2020 covering several areas of Earth. An example is shown in Figure 1. Care was taken to select scenes with minimal cloud cover or no cloud cover for the use of archived images. It was also ensured that at the time of image acquisition, a Sentinel-2 image was available within a week or less from that time. Finally, 46 large images were selected, covering various areas with urban, rural, agricultural, maritime, and other land uses. However, considering the structural differences in densely populated urban areas and the lack of images in these regions, they were excluded from the training data.
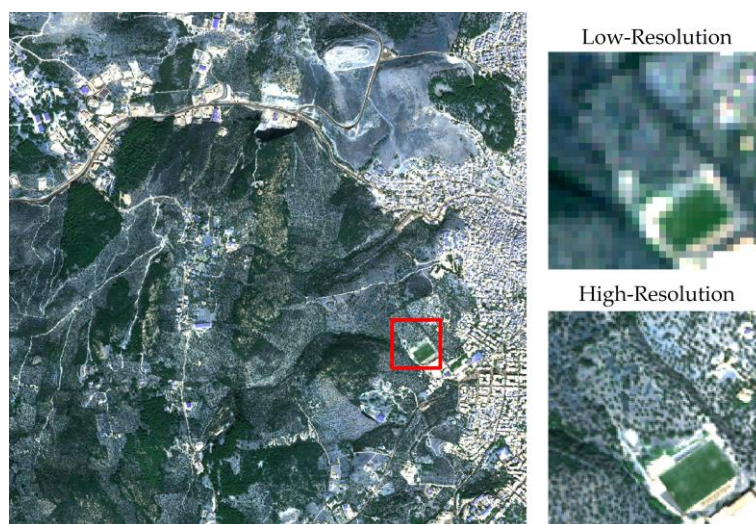

Figure 1. Sample of LR-HR pairs from the dataset.

After properly selecting the Sentinel-2 and Worldview image pairs, initial preprocessing steps were performed, such as converting the Digital Number (DN) to radiance. Additionally, the Sentinel-2 images were co-registered with their corresponding Worldview images, which have a higher spatial resolution. Then, the Worldview bands were resampled to a spatial resolution of 2.5 meters. Finally, the Sentinel-2 and Worldview images were respectively divided into 64×64 and 256×256 patches and then augmented based on the frequency of each land use.

2.2. Proposed Method

In this section, we describe our experiments conducted to verify our methods. The method of generating the used data set was explained in past section. Generally, 6500 patches were collected and segregated into two sets, comprising 90% for training and 10% for testing for the purpose of validation. Therefore, for a better comparison, both the ESRGAN network and the proposed method were trained once. For quantitative evaluation of the results, the well-known metrics of Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index (SSIM) were used, which are widely used in previous studies. Additionally, Spectral Angle Mapper (SAM) metrics were selected to assess the spectral quality.

The ESRGAN [10] network was presented in 2018 by Wang et al. The generator network architecture in the post-upsampling framework. In the beginning, a convolutional layer as a feature extractor is used to generate 64 features with a kernel of 3x3 dimensions and stride 1. Then, the generated features are input to a deeper feature extractor, which is composed of 23 Residual of Residual Dense Blocks (RRDBs). The RRDB blocks contain one residual architecture that uses three other residual architectures. Finally, after the upsampling layers, the extracted features are combined by two convolution layers and the

output of the network is obtained in the form of three features that it represents three RGB spectral bands.

In our proposed method, by extracting spatial features such as morphological profile and sharpening operation, we tried to reduce the pixelation in super-resolved images. Morphological analyses are designed using various shapes to extract information present in the form of spatial relationships between pixels. One common method in employing morphological analyses is to create a morphological profile. This means that different morphological operators are first applied to the image, and then their results are arranged alongside each other based on a specific pattern and utilized in a computational model [12]. In this study, we employed two morphological operators, Opening and Closing, to generate the morphological profile. For each of the three RGB spectral bands, these operators were applied three times, and their results were arranged sequentially as shown in Figure 2.
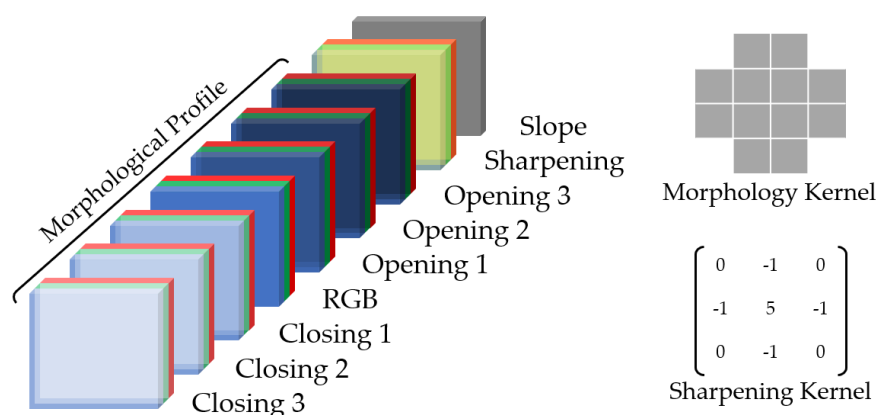

Figure 2. The order of computed features and their utilized elements.

Additionally, we used the slope to prevent the degradation of land surface topography by the model. For this purpose, corresponding to each data, the Shuttle Radar Topography Mission (SRTM) v3 Digital Elevation Model (DEM) was acquired, and the regional slope was calculated. The use of the slope is advantageous compared to the elevation model because it has a threshold and can be applied without considering the image's position on the Earth.

After preparing the required input data, fine-tuning and training the implemented model become crucial. In GAN networks, both the Generator (G) and Discriminator (D) networks are trained simultaneously. These networks engage in an adversarial relationship, striving to enhance each other's performance. Thus, maintaining equitable competition between them is of paramount importance. Therefore, initially, the G network was trained independently, and then the trained network was pitted against the D network.

3. Result

To compare our proposed method and ESRGAN network, the three metrics were considered PSNR, SSIM, and SAM that the outcomes shown in the Table 1.

Table 1. Comparison of quantitative evaluation metrics on the test dataset between the ESRGAN network and the proposed method. The entered values in the table represent the calculated mean and standard deviation ($\mu \pm \sigma$).

Metric		ESRGAN	Proposed Method
PSNR (dB)	$\uparrow_0^\infty$	37.34 ± 3.95	37.23 ± 4.23
SSIM	$\uparrow_{-1}^{+1}$	0.90 ± 0.05	0.92 ± 0.05
SAM (radian)	$\downarrow_0^\pi$	0.12 ± 0.06	0.10 ± 0.06

According to the table 1, provided method could achieve the incredible result at SAM and SSIM metrics, 0.10 radian and 0.92 respectively that were better than ESRGAN network with 0.12 radian in SSIM and 0.90 in the SAM. By contrast, at the PSNR metric ESRGAN gain higher values with 37.34 dB and the approach provided shows the 37.23 dB.

On the other hand, to compare the images visualize, the outcome the real, approach provided, and ESRGAN network shown in Figure 3, Moreover, in order to survey and evaluation the images spectral the RGB bands histogram were considered in Figure 4.
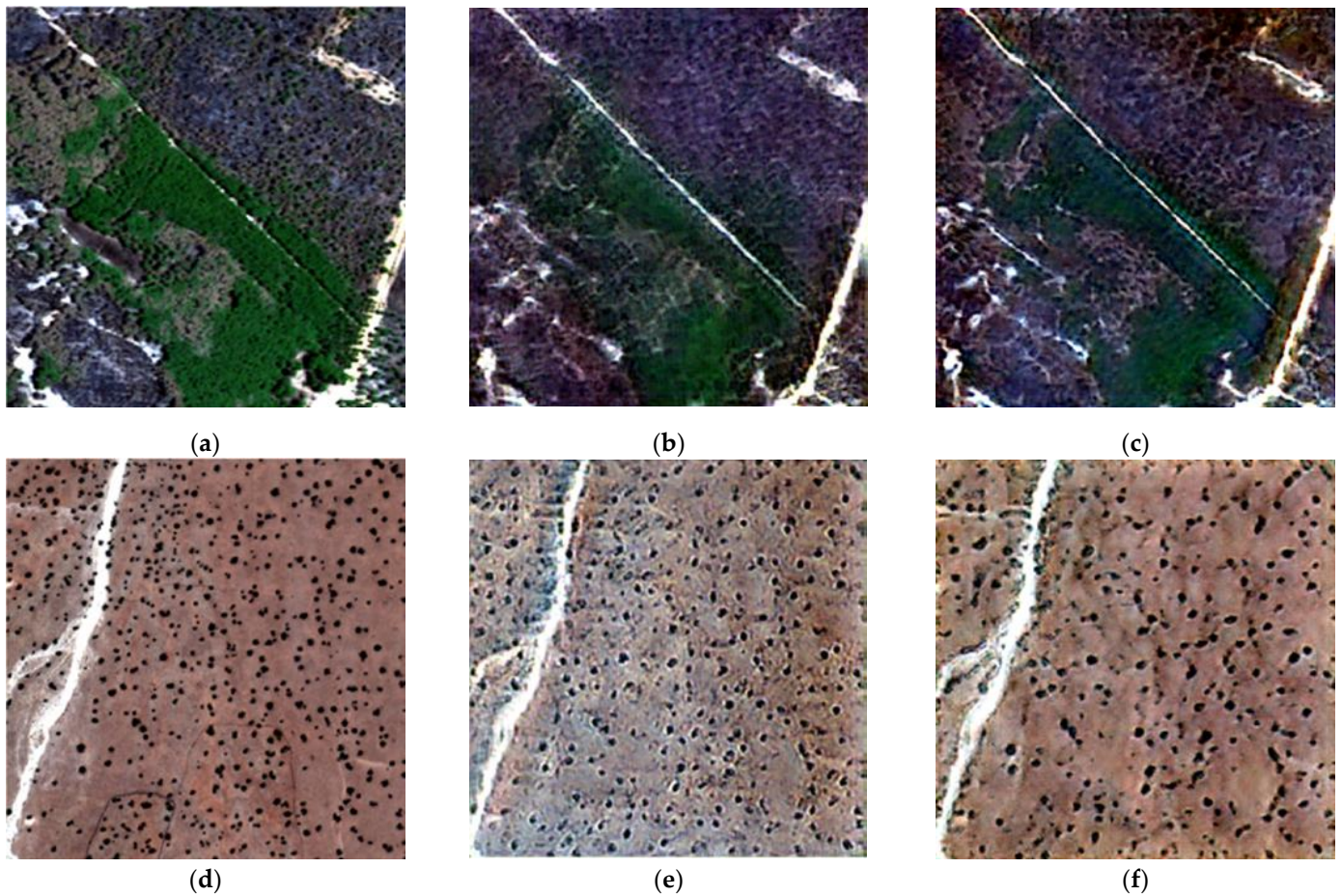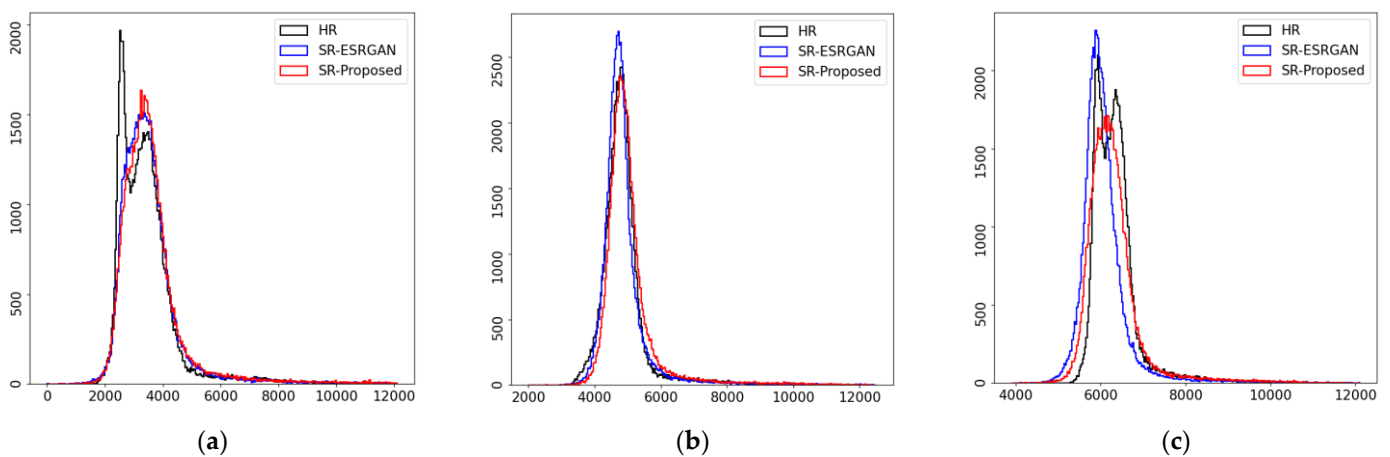

Figure 3. Visual comparison (RGB) of Worldview (**a**) and (**d**), Super-resolved by proposed model (**b**) and (**e**), and Super-resolved by ESRGAN (**c**) and (**f**) images; the values of PSNR, SSIM and SAM are given, respectively: (**b**) 38.07 dB/0.93/0.16 rad., (**c**) 37.89 dB/0.93/0.16 rad, (**e**) 34.80 dB/0.84/0.12 rad., (**f**) 37.91 dB/0.91/0.09 rad.
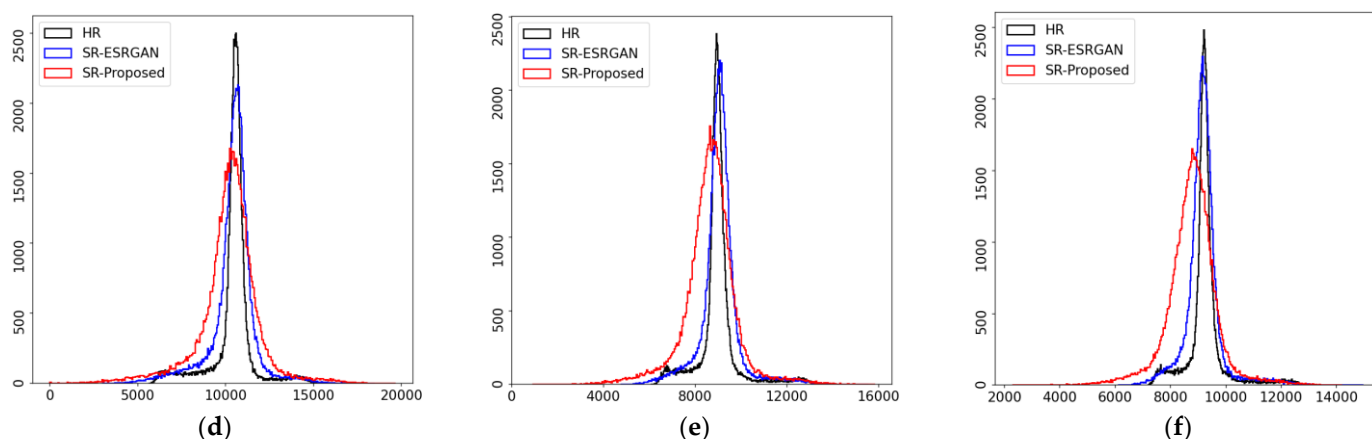
Figure 4. Comparison between the histograms of the Red (**a**) and (**d**), Green (**b**) and (**e**), and Blue (**c**) and (**f**) bands from the images in Figure 3. In each histogram, black, red, and blue colors correspond to worldview, Super-resolved by the proposed model, and Super-resolved by ESRGAN images, respectively.

4. Conclusion

In this paper, the susceptibility of satellite images, such as Sentinel-2, to pixelation super-resolved images was addressed. The proposed method in this study involved utilizing spatial features in the model input. Implementing this approach managed to enhance the network's performance and improve the visual results. Therefore, investigating and focusing on the selection or extraction of appropriate features is one of the important considerations.

References

1. L. Salgueiro Romero, J. Marcello, and V. Vilaplana, "Super-resolution of sentinel-2 imagery using generative adversarial networks," *Remote Sens.*, vol. 12, no. 15, p. 2424, 2020.
2. G. Rohith and L. S. Kumar, "Paradigm shifts in super-resolution techniques for remote sensing applications," *Vis. Comput.*, vol. 37, no. 7, pp. 1965–2008, 2021.
3. S. E. Armannsson, M. O. Ulfarsson, J. Sigurdsson, H. V. Nguyen, and J. R. Sveinsson, "A comparison of optimized Sentinel-2 super-resolution methods using Wald's protocol and Bayesian optimization," *Remote Sens.*, vol. 13, no. 11, p. 2192, 2021.
4. M. Gargiulo, A. Mazza, R. Gaetano, G. Ruello, and G. Scarpa, "Fast super-resolution of 20 m Sentinel-2 bands using convolutional neural networks," *Remote Sens.*, vol. 11, no. 22, p. 2635, 2019.
5. L. Liebel and M. Körner, "Single-image super resolution for multispectral remote sensing data using convolutional neural networks," *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.*, vol. 41, pp. 883–890, 2016.
6. J. M. Haut, R. Fernandez-Beltran, M. E. Paoletti, J. Plaza, and A. Plaza, "Remote sensing image superresolution using deep residual channel attention," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 11, pp. 9277–9289, 2019.
7. M. Galar Idoate, R. Sesma Redín, C. Ayala Lauroba, L. Albizua, and C. Aranda, "Learning super-resolution for Sentinel-2 images with real ground truth data from a reference satellite," *ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci. V-1-2020 9-16*, 2020.
8. 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, 2015.
9. I. Goodfellow *et al.*, "Generative adversarial nets," *Adv. Neural Inf. Process. Syst.*, vol. 27, 2014.
10. X. Wang *et al.*, "Esrgan: Enhanced super-resolution generative adversarial networks," in *Proceedings of the European conference on computer vision (ECCV) workshops*, 2018, pp. 0–0.
11. T. J. Sejnowski, "The unreasonable effectiveness of deep learning in artificial intelligence," *Proc. Natl. Acad. Sci.*, vol. 117, no. 48, pp. 30033–30038, 2020.
12. M. Fauvel, J. A. Benediktsson, J. Chanussot, and J. R. Sveinsson, "Spectral and spatial classification of hyperspectral data using SVMs and morphological profiles," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 11, pp. 3804–3814, 2008.