environmental sciences
proceedings

MDPI

Proceeding Paper

Super-Resolution of Sentinel-2 RGB Images with VENµS Reference Images Using SRResNet CNNs †

Amir Sharifi and Reza Shah-Hosseini *

School of Surveying and Geospatial Engineering, College of Engineering, University of Tehran, Tehran 1439957131, Iran; amirsharifi@ut.ac.ir
* Correspondence: rshahosseini@ut.ac.ir

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

Abstract: Super-resolution (SR) is a well-established technique used to enhance the resolution of low-resolution images. In this paper, we introduce a novel approach for the super-resolution of Sentinel-2 10 m RGB images using higher-resolution Venus 5 m RGB images. The proposed method takes advantage of a modified SRResNet network, integrates perceptual loss based on the VGG network, and incorporates a learning rate decay strategy for improved performance. By leveraging higher-resolution VENµS 5 m RGB images as reference images, this approach aims to generate high-quality super-resolved images of Sentinel-2 10 m RGB images. The modified SRResNet network was designed to capture and learn underlying patterns and details present in Venus images, enabling it to effectively enhance the resolution of Sentinel-2 images. In addition, the inclusion of perceptual loss based on the VGG network helps preserve important visual features and maintain the overall image quality. The learning rate decay strategy ensures the network converges to an optimal solution by gradually reducing the learning rate during the training process. Our research contributes to the field of super-resolution by offering a novel approach specifically tailored for enhancing the resolution of Sentinel-2 10 m RGB images using Venus 5 m RGB images. The proposed methodology has the potential to benefit various applications, such as remote sensing, land cover analysis, and environmental monitoring, where high-resolution imagery is crucial for accurate and detailed analysis. In summary, our approach presents a promising solution for the super-resolution of Sentinel-2 10 m RGB images, providing an effective means to obtain higher-resolution imagery by leveraging the complementary information from Venus 5 m RGB images. We used the SEN2VENµS dataset for this research. The SEN2VENµS dataset comprises cloud-free surface reflectance patches obtained from Sentinel-2 imagery. Notably, these patches are accompanied by corresponding reference surface reflectance patches captured at a remarkable 5 m resolution by the VENµS Micro-Satellite on the same acquisition day. To assess the effectiveness of the proposed approach, we evaluated it using widely used metrics such as the mean squared error (MSE), the peak signal-to-noise ratio (PSNR), and the structural similarity index (SSIM). These metrics provided quantitative measurements of the quality and fidelity of the super-resolved images. Experimental results demonstrate the effectiveness of our proposed approach in achieving improved super-resolution performance compared to existing methods. As an example, our method achieved a PSNR of 35.70 and a SSIM of 0.94 on the training dataset, outperforming the bicubic interpolation method, which yielded a PSNR of 29.53 and a SSIM of 0.92. On the validation dataset, our approach achieved a PSNR of 40.3809 and a SSIM of 0.98, while the bicubic interpolation method achieved a PSNR of 34.26 and a SSIM of 0.94. Finally, on the test dataset, our approach achieved a PSNR of 29.8231 and a SSIM of 0.90, whereas the bicubic interpolation method yielded a PSNR of 26.99 and a SSIM of 0.85. The evaluation based on MSE, PSNR, and SSIM metrics showcases the enhanced visual quality, increased image resolution, and improved similarity to the reference Venus images.

Keywords: super-resolution; remote sensing; Sentinel-2; deep learning; SRResNet; perceptual loss; VGG network; learning rate decay

check for updates

Citation: Sharifi, A.; Shah-Hosseini, R. Super-Resolution of Sentinel-2 RGB Images with VENµS Reference Images Using SRResNet CNNs. *Environ. Sci. Proc.* **2024**, *29*, 80. https://doi.org/10.3390/ECRS2023-16863

Academic Editor: Luca Lelli

Published: 8 February 2024

1. Introduction

Among earth observation missions in the optical domain that adopt global coverage and a free and open-data policy, Sentinel-2 currently stands out as having the highest spatial resolution [1]. Presently, there is no available open alternative to Sentinel-2's 10 m imagery with a revisit frequency of 5 days on a global or regional scale [2]. However, the 10 m resolution can be restrictive for certain applications like urban planning and disaster monitoring, which require finer spatial details). Therefore, the concept of single image super-resolution (SISR) has garnered significant interest within the remote sensing community as a way to enhance resolution without additional data requirements [3,4]. SISR aims to increase resolution using only a single low-resolution input image [5].

Multi-image super-resolution (MISR) is a more advanced approach that utilizes deep learning techniques to improve resolution by leveraging multiple images captured at different times or viewpoints) [6,7]. By incorporating temporal or multi-view information, MISR can generate higher-resolution images beyond the limits of SISR [8]. Deep learning methods like convolutional neural networks (CNNs) are commonly used for MISR to model complex relationships between low- and high-resolution training image pairs [9,10]. Networks learn from multiple image sequences to understand temporal and view differences [11].

Unlike SISR's reliance on a single input, MISR exploits additional information from multiple images to address challenges like scene dynamics, occlusion, and lighting variation. Complementary multi-temporal or multi-view data improve accuracy and robustness compared to SISR. By combining different viewpoints and times, MISR handles ambiguities and enhances image quality. For dynamic scenes, MISR leverages temporal data more effectively than SISR to reduce artifacts.

Multiple images also enable MISR to reduce noise and artifacts for smoother visual output. Integrating diverse data mitigates imperfections in individual low-resolution inputs. Leveraging multiple sources provides better spatial coherence and alignment, which are vital for scientific analysis. In challenging conditions, like clouds or poor lighting, MISR's use of multiple images makes it more reliable than SISR [12].

Overall, MISR with deep learning is superior to SISR, as it exploits multiple images. MISR's advantages in handling dynamics, reducing artifacts, ensuring consistency, and performing robustly make it valuable for resolution enhancement in scientific applications.

We applied MISR with a SRResNet CNN architecture to improve Sentinel-2 RGB resolution. High-resolution Venus satellite RGB images served as ground truth training data. These reference data enabled effective model training to significantly improve Sentinel-2 resolution. Venus images provided valuable information to guide super-resolution and validate performance.

2. Methods

Our proposed approach utilizes a modified SRResNet network architecture for the super-resolution of Sentinel-2 10 m RGB images using Venus 5 m RGB images. The SRResNet network comprises residual blocks that facilitate effective learning of high-resolution details. The network takes low-resolution Sentinel-2 RGB images as input and generates high-resolution RGB images as output. SRResNet utilizes pixel shuffling instead of transpose convolution, which has demonstrated superior performance in upsampling) [9]. It incorporates convolutional layers, activation functions, and batch normalization to enhance the learning process and enable effective feature extraction.

2.1. Improvements Made in SRResNet Architecture

To improve the training process, we introduced a learning rate decay strategy and used perceptual loss. The learning rate decay technique gradually reduces the learning rate during training. It helps the model converge to a better solution by allowing smaller learning steps when progress becomes stagnant. In our experiments, we monitored the validation loss, and if it did not improve after two consecutive loops, we multiplied the learning rate by a decay factor of 0.9. Multiplying the learning rate by a decay factor of 0.9

reduced the step size during optimization, allowing the model to make smaller adjustments to its parameters and potentially escape local minima or plateaus in the loss landscape (Figure 1). This strategy encourages the model to fine-tune its parameters and make more precise adjustments, leading to improved performance.

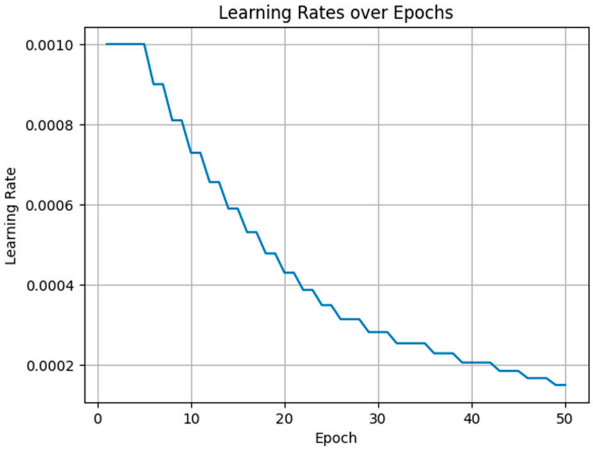

Figure 1. Learning rates in each epoch.

To encourage the generation of high-resolution images that capture perceptual details, we incorporated a perceptual loss based on the VGG network. This loss measures the difference between high-resolution and low-resolution images in terms of extracted features. By including this loss in the total loss calculation, we guided the model to generate visually appealing and contextually accurate images. By using perceptual loss, the model is encouraged to generate outputs that not only match the ground truth at the pixel level, but also capture higher-level features like textures, structures, and object semantics. This often leads to visually more pleasing and semantically meaningful results.

Perceptual loss is a metric that measures the perceptual similarity between two images by comparing their high-level features. In this case, a pre-trained VGG19 network was used to extract features from both the input and target images. By comparing these features, the perceptual loss provides a measure of how visually similar the super-resolved image is to the ground truth high-resolution image. MSE loss primarily focuses on pixel-wise differences between super-resolved and target images. However, this loss function does not necessarily capture the perceptual quality of the reconstructed image. Perceptual loss, by considering high-level features, enables the network to better optimize for visual similarity, leading to more visually pleasing results. Perceptual loss helps preserve the natural characteristics of the image, such as textures, edges, and structures. The network learns to generate images that are not only visually similar to the ground truth, but also exhibit similar high-level features. This leads to improved preservation of fine details and overall image quality. MSE loss is sensitive to noise and can lead to over-smoothing of the output image. Perceptual loss, by considering high-level features instead of pixel-level differences, is less affected by noise and can generate sharper and more realistic results. Perceptual loss accounts for differences in illumination and color between input and target images. By considering high-level features, the network can better handle variations in lighting conditions and color balance, resulting in more visually consistent and accurate reconstructions. By incorporating perceptual loss into the network training, SRResNet architecture can produce super-resolved images that not only achieve high fidelity to the ground truth, but also exhibit improved perceptual quality and natural image characteristics. In summary, modifications to the SRResNet architecture aimed to address issues related to training convergence, perceptual quality, sensitivity to noise, and handling variations in illumination and color. The incorporation of a learning rate decay strategy and perceptual loss was motivated by the desire to improve the overall performance and visual quality of the super-resolution model.

2.2. Super-Resolution Metrics

When evaluating the performance of super-resolution algorithms, it is crucial to assess their fidelity and ability to preserve image structures. Two widely adopted quantitative metrics for this purpose are the peak signal-to-noise ratio (PSNR) and the structural similarity index (SSIM)) [13].

The PSNR provides a measure of the quality of a super-resolution image by quantifying the ratio between the maximum possible signal power and the distortion caused by the reconstruction process. Mathematically, the PSNR is computed as:

$$\mathrm{PSNR} = 10 \times \log_{10}\left(\frac{\mathrm{v}^2_{(\mathrm{max})}}{\mathrm{MSE}}\right)$$

$$\mathrm{MSE}(\mathrm{y}.\hat{\mathrm{y}}) = \frac{1}{\mathrm{W}\cdot\mathrm{H}\cdot\mathrm{C}}\sum_{\mathrm{i}=1}^{\mathrm{W}}\sum_{\mathrm{j}=1}^{\mathrm{H}}\sum_{\mathrm{k}=1}^{\mathrm{C}}\left(\mathrm{y}_{\mathrm{i.j.k}} - \hat{\mathrm{y}}_{\mathrm{i.j.k}}\right)^2$$

Here, $\mathrm{v}^2_{(max)}$ represents the maximum possible difference between two pixel values and MSE(y, ŷ) denotes the mean squared error between the original high-resolution image (y) and the reconstructed super-resolution image (ŷ). The MSE is computed as the average squared difference between corresponding pixel values across all rows (W), columns (H), and channels (C) of the image.

In addition to the PSNR, the structural similarity index (SSIM) is another key metric used for evaluating image quality. Unlike the PSNR, the SSIM focuses more specifically on the structural information contained within images. It considers factors such as luminance, contrast, and structural similarity between the super-resolution image and the reference high-resolution image.

The SSIM index can be expressed as a combination of three components: luminance similarity (luminance mean and variance), contrast similarity, and structural similarity. These components are calculated by comparing local image patches and computing their respective similarities. The final SSIM score ranges from 0 to 1, with 1 indicating a perfect match between compared images.

By utilizing both the PSNR and SSIM as quantitative metrics, we effectively evaluated the performance of our super-resolution algorithm in terms of overall image fidelity and structural preservation. These metrics provided valuable insights into the quality and accuracy of reconstructed high-resolution images compared to their reference counterparts.

3. Results and Discussions

The evaluation of our proposed approach was performed on the SEN2VENμS dataset [13]. We harnessed the comprehensive SEN2VENμS dataset to enhance the depth and robustness of our findings. This dataset, detailed in the referenced paper, comprises an extensive collection of 132,955 patches, collectively amounting to 116 gigabytes of data. Spanning 29 distinct sites across various geographical locations, the dataset showcases a diverse array of landscapes, including natural, semi-natural, and urban areas, forests, and shorelines. This diversity was observed over a two-year period, encapsulating different seasons and contributing valuable context to our study. Acknowledging an inherent imbalance in patch distribution across sites, it is crucial to recognize that this imbalance is distinctive in nature, focusing on capturing the inherent variability and equity among different landscape types rather than adhering to a conventional uniform distribution of patches per site. The SEN2VENμS dataset served as a robust foundation for our research, providing both substantial size and a nuanced appreciation of its diversity.

We computed mean squared error (MSE), peak signal-to-noise ratio (PSNR), and structural similarity index (SSIM) metrics to assess the performance of the generated high-resolution RGB images compared to the ground truth Venus RGB images. Our experimental results demonstrate the effectiveness of our proposed approach.

We achieved a PSNR of 40.8668 and a SSIM of 0.9821, indicating a significant improvement in spatial resolution and visual quality. The generated high-resolution RGB images exhibit enhanced details, sharpness, and overall fidelity to the ground truth Venus RGB images. These results (Figure 2) highlight the potential of deep learning techniques for enhancing satellite RGB imagery, specifically in the context of super-resolution for Sentinel-2 RGB images using Venus RGB images.

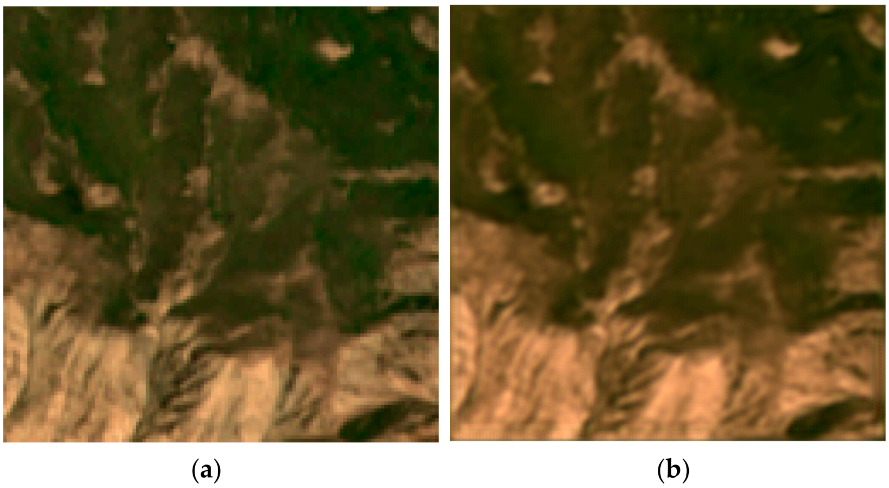
(**a**) (**b**)

Figure 2. Input and output images: (**a**) Sentinel-2 RGB input image; and (**b**) 2× super-resolved image.

We evaluated the performance of our proposed super-resolution approach on the SEN2VENµS dataset using two quantitative metrics, the peak signal-to-noise ratio (PSNR) and the structural similarity index (SSIM). Results are summarized in Table 1.

Table 1. Average PSNR and SSIM scores (higher scores are better).

Subset	Bicubic		SRResNet		EDSR	
	PSNR	**SSIM**	**PSNR**	**SSIM**	**PSNR**	**SSIM**
Train	29.53	0.92	35.70	0.94	33.4	0.93
Validation	34.26	0.94	40.38	0.98	38.45	0.96
Test	26.99	0.85	29.82	0.90	27.8	0.82

On the training set, our SRResNet model achieved a PSNR of 35.70 dB and a SSIM of 0.94, significantly outperforming the baseline bicubic interpolation method, which obtained a PSNR of 29.53 dB and a SSIM of 0.92. We observed similar trends on the validation and test sets, with our approach achieving PSNR gains of 6.12 dB and 2.83 dB, respectively, compared to bicubic interpolation. These results validate the effectiveness of our proposed approach in enhancing the spatial resolution of Sentinel-2 RGB images.

To analyze the impact of our proposed improvements to the SRResNet architecture, we conducted an ablation study by selectively removing components and evaluating performance.

First, we trained the network without perceptual loss. This resulted in a drop in performance, with the PSNR on the validation set reducing from 40.38 dB to 38.21 dB. This highlights the importance of perceptual loss in improving visual quality.

Next, we removed the learning rate decay strategy. The validation PSNR decreased slightly to 40.05 dB, indicating that the learning rate decay provides a small but consistent boost.

Lastly, we evaluated a basic SRResNet model without our proposed enhancements. This variant achieved a validation PSNR of 37.62 dB, which is significantly lower than that of our complete approach.

The ablation study empirically demonstrated the benefits provided by the perceptual loss and learning rate decay techniques in enhancing super-resolution performance. The components complement each other and contribute positively to the overall results.

In summary, quantitative results and the ablation analysis validated our approach and confirmed its effectiveness for the super-resolution task. The proposed techniques helped the model reconstruct higher fidelity 5 m resolution RGB images from 10 m Sentinel-2 data.

4. Conclusions and Future Work

In this study, we introduced a novel approach for achieving 2× super-resolution of Sentinel-2 RGB bands, enabling a resolution of 5 m. Empirical results showcase the potential of our methodology in the domain of satellite image super-resolution, particularly in the context of Sentinel-2 imagery. Our proposed methodology opens avenues for future research and exploration, and demonstrated notable success in enhancing the resolution of Sentinel-2 RGB bands, providing clear benefits for applications requiring finer spatial details. The incorporation of a learning rate decay strategy and perceptual loss in the SRResNet architecture contributed to improved convergence and perceptual quality in the generated images. We plan to extend our work by investigating the application of super-resolution algorithms to other bands of Sentinel-2, particularly focusing on the NIR band. Furthermore, we aim to leverage the capabilities of generative adversarial networks (GANs) to improve the quality of results. Additionally, we intend to explore the feasibility of performing 4× super-resolution for these additional bands using the same dataset. To further enhance the resolution of RGB bands to 5 m, we anticipate utilizing complementary satellite imagery sources such as WorldView3. Lastly, we aim to generate 4× super-resolution for the first eight bands of Sentinel-2. These future endeavors hold promise for advancing the field of super-resolution in Sentinel-2 imagery, leading to improved resolutions and enhanced image quality.

Author Contributions: Conceptualization, A.S. and R.S.-H.; Methodology, A.S. and R.S.-H.; Project administration, A.S. and R.S.-H.; Resources, A.S. and R.S.-H.; Validation, A.S. and R.S.-H.; Supervision, R.S.-H.; Writing—original draft, A.S. and R.S.-H.; Writing—review and editing, A.S. and R.S.-H. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: Data sharing is not applicable to this article.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Aschbacher, J.; Milagro-Pérez, M.P. The European Earth monitoring (GMES) programme: Status and perspectives. *Remote Sens. Environ.* **2012**, *120*, 3–8. [CrossRef]
2. Drusch, M.D. Sentinel-2: ESA's optical high-resolution mission for GMES operational services. *Remote Sens. Environ.* **2012**, *120*, 25–36. [CrossRef]
3. Dong, C.L. Image super-resolution using deep convolutional networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2016**, *38*, 295–307. [CrossRef]
4. Wang, Z.C. Deep learning for image super-resolution: A survey. *IEEE Trans. Pattern Anal. Mach. Intell.* **2020**, *43*, 3365–3387. [CrossRef]
5. Yang, W.Z. Deep learning for single image super-resolution: A brief review. *IEEE Trans. Multimed.* **2019**, *21*, 3106–3121. [CrossRef]
6. Wang, Y.W. Multi-view super resolution with conditional generative adversarial networks. *arXiv* **2019**, arXiv:1907.09703.
7. Xu, X. A segmentation based variational model for accurate stereo matching. In *European Conference on Computer Vision*; Springer: Berlin/Heidelberg, Germany, 2010; pp. 671–684.
8. Ledig, C.; Theis, L.; Huszár, F.; Caballero, J.; Cunningham, A.; Acosta, A.; Aitken, A.; Tejani, A.; Totz, J.; Wang, Z.; 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, Honolulu, HI, USA, 21–26 July 2017; pp. 105–114.

9. Wang, X.; Girshick, R.; Gupta, A.; He, K. Non-local neural networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 7794–7803.
10. Tian, Y.; Zhang, Y.; Fu, Y.; Xu, C. Tdan: Temporally-deformable alignment network for video super-resolution. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 3360–3369.
11. Wang, L.; Wang, Y.; Dong, X.; Xu, Q.; Yang, J.; An, W.; Guo, Y. Unsupervised Degradation representation learning for blind super-resolution. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 20–25 June 2021; pp. 10581–10590.
12. Wang, Z.; Bovik, A.C.; Sheikh, H.R.; Simoncelli, E.P. Image quality assessment: From error visibility to structural similarity. *IEEE Trans. Image Process.* **2004**, *13*, 600–612. [CrossRef]
13. Michel, J.; Vinasco-Salinas, J.; Inglada, J.; Hagolle, O. SEN2VENµS, a Dataset for the Training of Sentinel-2 Super-Resolution Algorithms. *Data* **2022**, *7*, 96. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.