



Real-ESRGAN: A deep learning approach for general image restoration and its application to aerial images

Şükrü Burak Çetin *¹

¹Istanbul Metropolitan Municipality, Directorate of Geographic Information Systems, Istanbul, Türkiye; sukruburak.cetin@ibb.gov.tr

Cite this study: Çetin, Ş. B. (2023). Real-ESRGAN: A deep learning approach for general image restoration and its application to aerial images. *Advanced Remote Sensing*, 3 (2), 90-99

Keywords

Image Processing
Artificial Intelligence
Aerial Imagery
GANs
Real-ESRGAN

Research Article

Received:20.06.2023
Revised:14.11.2023
Accepted: 05.12.2023
Published:10.12.2023



Abstract

General image restoration is a challenging task in computer vision, especially for images with complex scenes and noise. Practical algorithms for general image restoration, such as Enhanced Super-Resolution Generative Adversarial Networks (ESRGAN), have been developed to address this problem. Real-ESRGAN is a deep learning-based image restoration model that uses a generative adversarial network (GAN) to produce high-resolution images from low-resolution inputs. In recent years, Real-ESRGAN has gained significant attention for its impressive image restoration results on various types of images, including aerial images. Aerial images have unique challenges, such as high noise, low contrast, and blur, which affect the quality of the images. ESRGAN has been applied successfully to restore these images and enhance their visual quality, enabling better interpretation and analysis. In this article, we review the practical algorithms for general image restoration, with a focus on Real-ESRGAN and its application on aerial images. We discuss the architecture and application strategies of Real-ESRGAN, as well as its advantages and limitations. We also present examples of how Real-ESRGAN has been used in various applications, such as Segment Anything Model (SAM) and its application as object detection, classification, and segmentation. This study utilized the GÖKTÜRK II 2022 Istanbul Aerial Image dataset, which comprises 917,252 image chips with a resolution of (512x512) and a 3-channel RGB format. In order to enhance the visual quality, the image chips were upscaled to a higher resolution of (1024x1024) using a 2x scaling factor, resulting in a fourfold increase in data size equivalent to 2.684 TB with the same compression ratio. This project shows the potential of Real-ESRGAN in handling large-scale and diverse datasets, as well as its ability to enhance the visual quality of aerial images for real world image restoration, which is essential in various fields such as agriculture, urban planning, and disaster management.

1. Introduction

As the use of satellite imagery grows increasingly widespread across numerous applications, research into spatial resolution and its impact on image quality has grown significantly. Though the spatial resolution provided by an optical sensor on a satellite is typically expressed as a nominal value that signifies the pixel's footprint, the actual resolution may differ due to various factors such as atmospheric and imaging conditions, off-nadir angle of the satellite, and artefacts caused by the satellite's operation or optics. Due to the high cost of enhancing the optical components of a sensor to obtain high-resolution (HR) images, software solutions have been sought to reduce expenses. To this end, several super-resolution (SR) algorithms have been created to improve spatial resolution and generate HR images from one or more low-resolution (LR) images. SR approaches have been applied across various fields such as satellite and aerial image processing [1-4], medical image processing [5], facial and fingerprint image enhancement [6], text image enhancement, and compressed images and video enhancement [7-9]. This chapter serves as an introduction to the study, focusing on the advancement of spatial resolution through

the utilization of contemporary deep learning (DL) methods. It outlines the primary objectives of the research and provides concise descriptions of the subsequent chapters. Traditional approaches to enhance the spatial resolution of optical satellite images primarily employ pan-sharpening techniques, which involve merging panchromatic and multispectral (MS) images to generate high-resolution (HR) color images. Fusion techniques such as Principal Component Analysis (PCA) [10], Intensity Hue Saturation (IHS) [11], and Wavelet Transform [12] are commonly employed in this context. However, several limitations exist with regard to geometric integration [13], including potential color distortions, the absence of a fully automated and consistent method across diverse datasets, and the operator's expertise in the fusion technique.

In the pursuit of achieving superior resolution capabilities while maintaining cost efficiency, a remarkable advancement has been made through the development of Super-Resolution (SR) algorithms. These algorithms offer the potential to enhance spatial resolution without necessitating any modifications to the sensor structure. By harnessing the power of SR techniques, the utilization of Low-Resolution (LR) satellite images can be significantly expanded across a wide array of applications. This breakthrough has paved the way for novel possibilities and has opened doors to exploit LR imagery in unprecedented ways. SR methods are classified into two broad categories: frequency domain and spatial domain approaches [14]. Though frequency domain approaches, such as those described in [15-17], are computationally efficient, they are insufficiently effective at modeling complex problems. Almost all subsequent research on SR has been conducted in the spatial domain, despite the high computational cost. In the spatial domain, SR approaches are classified into two categories: single image SR (SISR) approaches and multi-image SR (MISR) approaches [18, 19]. SISR can make assumptions about the HR image based on a single input image, whereas MISR displays hidden HR details. MISR requires multiple LR images as input for the generation of HR images, although only one LR image is typically available. As a result, the use of SISR methods has grown in popularity. In recent years, there has been a surge of interest in methods based on convolutional neural networks (CNNs) and DL. Particularly, super-resolution studies based on Generative Adversarial Networks (GANs), such as [20-23], have become more favorable than traditional pan-sharpening methods. Because GAN-based approaches have the highest accuracy and visual performance, Real-Enhanced Super-Resolution Generative Adversarial Network (Real-ESRGAN) were preferred to enhance the image spatial resolution in this study. Real-ESRGAN, an enhanced version of ESRGAN, stands out as the chosen method, showcasing superior accuracy and visual performance, making it a more practical solution for real-world image restoration. It effectively addresses issues such as the removal of bothersome compression artifacts. Though the ground sampling distance (GSD) in GÖKTÜRK II Istanbul satellite image was reduced from 50 cm to 25 cm, the number of pixels in the image was increased and the clarity of the information captured in each pixel was clarified with the Real-ESRGAN algorithm. This study endeavors to contribute to the advancement of satellite image analysis by employing artificial intelligence (AI)-based super-resolution applications, specifically focusing on Enhanced Super-Resolution Generative Adversarial Networks (Real-ESRGAN). The preference for Real-ESRGAN in this context is rooted in its capability to deliver more realistic results that align with real-world examples. Real-ESRGAN's advanced features and training mechanisms make it particularly adept at enhancing satellite imagery in a manner that closely mirrors the nuances and details present in actual scenarios. By harnessing the power of Real-ESRGAN, this research endeavors to elevate the quality and authenticity of satellite image analysis, thereby addressing the demands for more accurate and lifelike representations in the field. The primary objective is to investigate the impact of enhanced image resolution quality and sharpness on the performance of prominent object detection and classification applications, with particular emphasis on the Segment Anything model. The overarching aim of this research is to assess whether the application of image enhancement techniques, such as super-resolution, can serve as a catalyst for improved object detection capabilities. In this context, the study seeks to determine whether these image enhancement applications could potentially play a pioneering role in bolstering the efficacy of object detection algorithms and if they can be integrated seamlessly as a pre-processing step in conventional object detection applications.

2. Generative Adversarial Networks (GANs)

The emulation of human cognition and creativity poses a formidable challenge for machines. Despite the intricacies involved in deciphering patterns from images and data, a task effortlessly performed by the human brain, the advent of innovative deep neural networks (DNNs) offers a promising avenue to overcome this hurdle. The realm of machine creativity becomes tangible through the lens of Generative Adversarial Networks (GANs). Originating in 2014, Ian Goodfellow introduced GANs as a distinctive class of DNNs [24]. Functioning in an unsupervised manner, GANs serve as neural network-based generative models, exhibiting the potential to create new, meaningful imagery and textual artworks.

The GAN framework comprises two integral subsystems known as the generator and discriminator. Operating in stark opposition to each other, these subsystems leverage Nash equilibrium from game theory to optimize the overall model. The generator, often likened to a "counterfeiter," endeavors to deceive the discriminator by producing synthetic data. Conversely, the discriminator, akin to a "detective," strives to accurately classify incoming data by discerning its authenticity. This adversarial interplay between the generator and discriminator

forms the core dynamic of GANs, encapsulating the essence of machine creativity and pattern recognition. Generative models operate within the realm of unsupervised learning, eliminating the need for explicit Y output variables when presented with X input data. The primary objective of generative models is to discern and internalize patterns inherent in the input data (X). Through this process, the generative model acquires the ability to extrapolate and generate novel outcomes, drawing upon the learned information encoded in the patterns identified within the input data. The discriminative model, characterized by its reliance on supervised learning, is commonly associated with classification tasks. Within this framework, the primary objective is to construct predictive models that assign labels to each input based on available data, predicting the specific class to which they belong. This methodology hinges on leveraging the richness of labeled data to discern and map relationships between inputs and their corresponding classes, facilitating accurate predictions in the classification process.

2.1. Super-Resolution Generative Adversarial Networks (SRGANs)

Super-Resolution Generative Adversarial Networks (SRGANs) are a class of generative models designed to address the challenge of enhancing image resolution. Introduced as a groundbreaking concept, SRGANs operate on the basis of a generator-discriminator architecture [25].

A succinct overview of SRGANs reveals the following steps: Input Low-Resolution Images: Provide low-resolution images as input to the generator. Generate Super-Resolution Images: The generator processes the input, producing high-resolution images as outputs. Discrimination Process: Subject the generated images to scrutiny by the discriminator, which evaluates their authenticity. Perceptual Enhancement: Incorporate a VGG net to introduce perceptual loss on a pixel-wise level, enhancing the sharpness of the generated synthetic images.

2.2. Enhanced Super-Resolution Generative Adversarial Network (ESRGAN)

ESRGAN thus represents a refined iteration in the evolution of super-resolution GANs (SRGANs), focusing on optimizing training efficiency and reducing complexity while maintaining the core principles of image super-resolution through adversarial learning. Enhanced Super-Resolution Generative Adversarial Networks (ESRGANs) introduce notable advancements and optimizations to the established framework of SRGANs. The key updates in ESRGANs encompass improvements in both the generator and discriminator components. On the generator enhancements side, Removal of Batch-Normalization Layers: ESRGANs eliminate the use of batch-normalization (BN) layers in the generator architecture, departing from the prevalent approach in SRGANs. This strategic choice results in heightened performance and reduced computational complexity. Residual in Residual Dense Block: ESRGANs employ a Residual in Residual Dense Block, an evolution beyond the standard residual block (Figure 1). This unique structure allows the outputs of all layers within a block to be transmitted to subsequent layers. By providing the model with a multitude of features to choose from, it enhances the model's ability to discern and prioritize relevant information. Additionally, ESRGAN incorporates Residual Scaling to scale down the residual outputs, preventing potential instability.

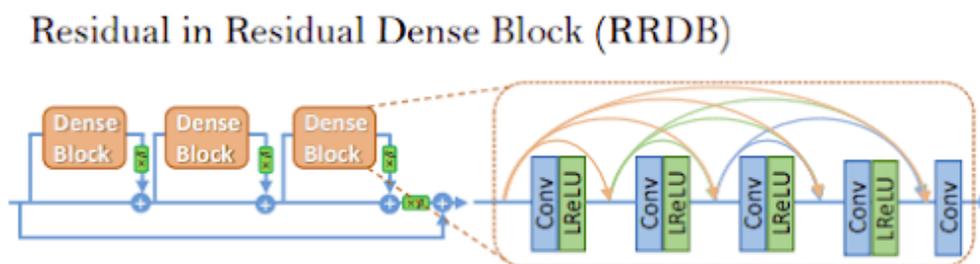


Figure 1. Residual in residual dense block.

On the discriminator updates side, Relativistic Loss: A significant addition to the discriminator is the inclusion of relativistic loss. This novel loss function estimates the probability of a real image being relatively more realistic than a predicted fake one. By incorporating relativistic loss, the model is incentivized to continually improve its realism in comparison to fake images. Perceptual Loss Modification: The perceptual loss in ESRGANs undergoes a slight modification. Unlike SRGANs, where the loss is based on features after the activation function, ESRGANs shift the focus to features right before the activation function. This adjustment aims to refine the perceptual quality of the generated images. Comprehensive Loss Function: The total loss in ESRGANs is a composite of the GAN loss, perceptual loss, and the pixel-wise distance between the ground truth high-resolution images and the generated counterparts. This comprehensive loss function encapsulates multiple aspects, ensuring a holistic optimization approach during training. In summary, ESRGANs represent a refined iteration of super-resolution GANs, introducing architectural adjustments and loss function enhancements to elevate both performance and image quality.

2.3. Real-Enhanced Super-Resolution Generative Adversarial Network (Real-ESRGAN)

Real-ESRGAN, an enhanced version of ESRGAN, represents a more practical solution for real-world image restoration by effectively addressing issues such as the removal of bothersome compression artifacts.

3. Material and Method

The GÖKTÜRK II satellite had been supplied by General Directorate of Mapping, consisted of 18 parts, whereas the Pleiades satellite consisted of 4 parts, making up a total of 22 sheets with each image approximately ~50 GB in size. The sheet images were initially received as Raw 2 images, and it was found that the metadata information was not preserved after processing with the Real-ESRGAN algorithm. The images were re-obtained after being orthorectified. To make the application effective on the image, it was started with a single computer. The Real-ESRGAN algorithm was optimized to work on the CPU. Processing the chip images with the CPU took approximately 14 seconds per chip, which delayed the project timeline. To improve application performance, the algorithm was optimized to use GPU tensors in PyTorch. PyTorch 1.12.1 GPU, CUDA 11.6, CUDNN 8.0 versions were used in this study. The processing power per chip was reduced from 14 seconds to 5 seconds. The algorithm files processed all the chip images individually from a single file. A multi-folder system was used, and thread optimization was used to process files in parallel. The processing performance per chip was reduced from 5 seconds to between 1 and 1.7 seconds, depending on the computer's GPU processing power, after optimization using GPU tensors. Despite using a high-performance GeForce GTX 3060 graphics card on a single computer, it was found that the application's chip processing productivity was not sufficient. To process all the chips in parallel form with similar and equal parts, 5 server computers were used. To be able to process in parallel, the distribution of files was linearly proportional to the computer's GPU processing power, shared through local sharing folders. Sharing ratios of 24%, 23%, 23%, 15%, and 15% were respectively used for 1 NVIDIA RTX 4000, 2 GeForce GTX 3060, and 2 mobile 1080 graphics card computers, according to the computer's power. After the chip images obtained from each map section were transferred to the shared folders at the mentioned ratios, the transfer of chip images from these shared folders to the local data areas of the server computer was ensured. Since processing each data one by one took too much time for a large number of chip images, monitoring the time was an important step, and time measurement mechanisms were created to show the processing capacity of the current files.

The processing of chip images involves three main stages, along with a complementary step (Figure 2), following the selection of an appropriate Real-ESRGAN model. These stages consist of pre-processing, where color space transformations are performed; main processing, where chip shape transformations are applied; and post-processing, where last touch transformations are carried out. In addition, complementary transformations are applied, including raster to mosaic transformation and symbology equalization. This study describes the methodology used to process chip images and explains the rationale behind each stage and complementary step.

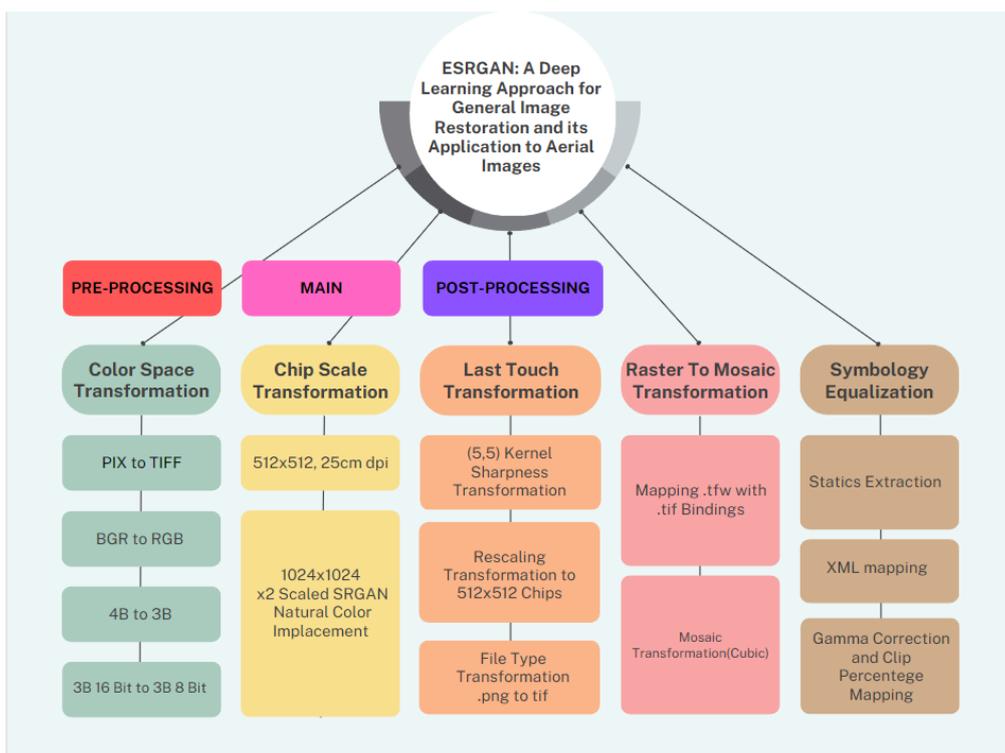


Figure 2. Experimental procedures.

3.1. The Experimental Procedures

In this study, a super-resolution model was investigated to achieve optimal image clarity by applying three models trained with disparate features to GÖKTÜRK II raw satellite image chip data sample (Figure 3). The models included RealESRGAN_x2plus (Figure 4), RealESRGAN_x4plus_netD Result (Figure 5), and RealESRNet_x4plus (Figure 6).

The impact of each model on the raw images was examined to determine the most effective in minimizing departure from realistic image flow, a handicap of image enhancement. RealESRGAN_x4plus (Figure 6) was found to be the model that deviated the least from realistic image flow and was therefore selected for the image enhancement process.



Figure 3. Raw Data Sample.



Figure 4. RealESRGAN_x2plus Result.



Figure 5. RealESRGAN_x4plus_netD Result.



Figure 6. RealESRGAN_x4plus Result.

3.1.1. Pre-Processing (Color Space Transformation)

The satellite images had been supplied by General Directorate of Mapping, including GÖKTÜRK II and Pleiades, were compressed in the .pix format. The image tiles were converted from the PIX format to the TIFF format. The raw images were provided in the BGR color space, and the image tiles with the BGR color space were transformed into the RGBA color space.

Since the trained models were created from 3D images, the transparency information for the image tiles was removed, and the images were reduced from four dimensions to three dimensions. As the raw images had a depth of 16 bits, each band was reduced from 16 bits to 8 bits to enable examination of the resulting images.

3.1.2. Chip Scale Transformation

Each image tile was converted into 512x512 resolution image chips using the Export Data for Deep Learning geoprocessing tool in ArcGIS Pro, with a cell size reduction from 50 cm to 25 cm. During this process, a total of 917,252 tiles were obtained from the Istanbul Satellite 2022 image. The coordinate information for each chip

image was stored in .tfw files. Each chip image was then subjected to x2 scale Natural Color Emplacement using the RealESRGAN_x4plus model and scaled to 1024x1024 dimensions with minimal loss.

To improve the effectiveness of the Real-ESRGAN model on the image, anti-aliasing technique was applied to the raw chip image. Anti-aliasing is a technique used in computer graphics to reduce the visual artifacts that can appear on images or text displayed on a screen. It works by smoothing out the jagged edges of diagonal or curved lines, making them appear smoother and more natural to the eye.

3.1.3. Post-Processing (Last Touch Transformation)

The noise in the chip images generated with the RealESRGAN_x4plus model at 1024x1024 resolution was minimized. As a result, morphological sharpening algorithms could be applied to the image without noise content. A sharpening kernel with a depth of (5, 5) from the OpenCV library was applied to the improved chip images as a post-processing method. Pastel-colored state of the image could occur though improving the image in Real-ESRGAN applications. To prevent this from being transferred to the image as much as possible, the image made the image expression in the edge regions of the satellite image more pronounced. Applying morphological sharpening algorithms to the scaled image ensures that the details stand out more smoothly. To be able to properly integrate the sharpened 1024x1024 chip images into the mosaic process, they were scaled back to 512x512 in the plane where their coordinates were stored. The model did not support .tiff format files. Therefore, the resulting .png files were converted back to the tiff format to prevent information loss in the pyramid image.

The chip images that were enhanced with the Real-ESRGAN model and morphologically sharpened were scaled back down to a resolution of 512x512 to ensure proper stitching during the mosaic process. Then, using the ArcGIS Pro "Mosaic to New Raster Geoprocessing Tool" with the Cubic resampling method, the chip images were mosaicked. The mosaic operator was set to "Last" chip tile and the mosaic colormap mode was set to "First" chip tile to create the mosaic image. The symbology and stretch type information obtained from the raw imagery were stored in .xml format for each tile image, and then applied to the enhanced mosaic image with statistical information. A Gamma correction value of 0.7 was stored as the stretch type, while a Clip Percentage Mapping value of 0.250 was assigned for each band.

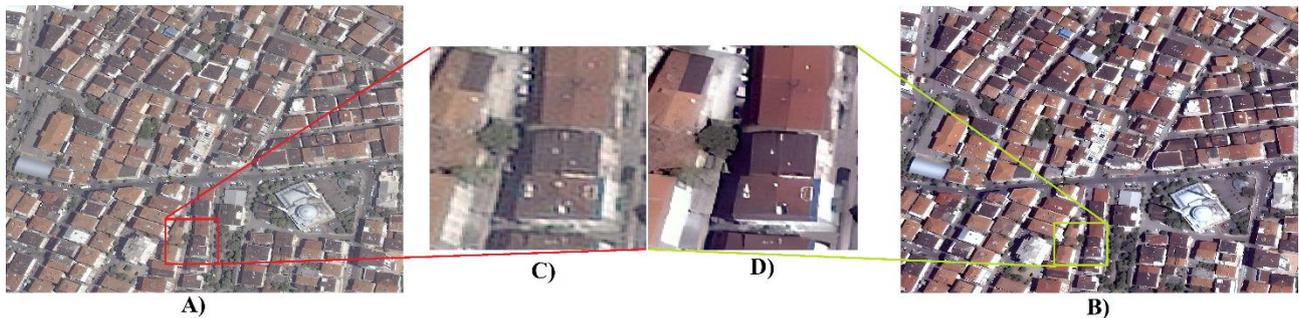


Figure 7. A) 2022 Istanbul Aerial Raw Data, **B)** Real-ESRGAN-Enhanced Data, **C)** Raw Data Zoomed-In Part, **D)** Real-ESRGAN-Enhanced Data Zoomed-In Part.

4. Results

In this study, we aimed to investigate the impact of super resolution applied raster data versus raw data on the performance of an object detection model, specifically the Segment Anything Model (SAM). The Segment Anything Model (SAM) is an object detection model that is based on a state-of-the-art computer vision algorithm known as Mask R-CNN (Region-based Convolutional Neural Network) which combines the object detection and image segmentation capabilities [26]. In this study, we used segment-geospatial, an improved spatial data segmentation tool that uses a segment anything model infrastructure. Segment-Geospatial is a powerful tool for geospatial data segmentation and classification, providing researchers and academics with a flexible and efficient framework for their geospatial analysis needs [27]. By leveraging the library's functionalities, users can gain valuable insights into their geospatial datasets and extract meaningful information for various academic applications. We compared the performance of SAM on two sets of data, one processed with super resolution and the other without any preprocessing, in terms of accuracy and efficiency. The object of interest in this study was rooftop detection, as it is a critical task for urban planning and management. SAM is designed to detect and segment multiple objects of different classes within an image. It can identify objects in complex scenes, even when they are partially obscured or occluded by other objects. SAM can be trained on various datasets to detect different types of objects such as vehicles, pedestrians, and buildings. The main advantage of SAM over traditional object detection models is its ability to produce pixel-level object masks. This makes it highly accurate in localizing the objects of interest and enables it to identify the boundaries of each object in the image. Additionally, SAM can be used to perform instance

segmentation, which is a more advanced form of object detection that assigns a unique label to each instance of an object in an image.

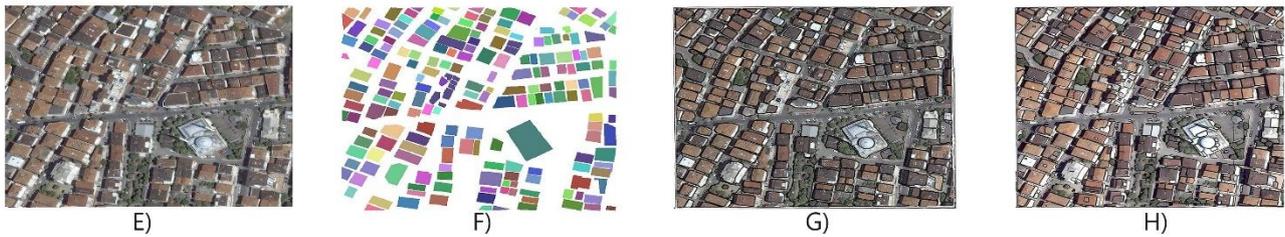


Figure 8. E) Original Sample, F) Ground-Truth Data, G) Raw Data SAM Model Output, H) ESRGAN-Enhanced Data SAM Model Output.

Table 1. Comparison of proposed method.

Sam Model	TP	FP	FN	Precision	Recall	F1-Score	IoU
Applied on Raw Data	197	7	5	0.965	0.975	0.970	≈ 0.942
Applied on Real-ESRGAN-Enhanced Data	207	24	2	0.896	0.991	0.941	≈ 0.888

The study aimed to evaluate the performance of autonomous SAM detection models in identifying rooftop vector polygons. The ground truth data consisted of 209 polygons, and SAM identified 197 of these correctly, with 7 false positive detections and 5 false negatives on Raw Istanbul Aerial Data. Moreover, SAM identified 207 of these correctly, with 24 false positive detections and 2 false negatives on Real-ESRGAN Applied Istanbul Aerial Data (Table 1) (Figure 8).

These results suggest that though both models were able to identify the majority of rooftops accurately (Figure 8), SAM results on Real-ESRGAN applied data had a higher false positive rate, which could result in additional resources being expended in verifying the accuracy of detected rooftops. Conversely, SAM results on raw data had a lower false positive rate, but also missed a significant number of rooftops, resulting in higher false negatives. Based on the findings presented, it is difficult to conclude that Real-ESRGAN-enhanced data provides better results than raw data. Though the Real-ESRGAN-enhanced data segmentation model achieved a higher Recall score than the raw data segmentation model, it also had a lower Precision score and a higher number of false positive predictions. The choice of which data to use ultimately depends on the specific application and the trade-offs between Precision and Recall that are acceptable for that application. If minimizing false positives is critical, the raw data segmentation model may be preferred. On the other hand, if identifying as many true positives as possible is the priority, the Real-ESRGAN-enhanced data segmentation model may be more suitable.

5. Discussion

Digitization of various types of physical records and materials has become increasingly important in recent years, as organizations seek to leverage digital technologies for better management, access, and analysis of their data. However, manual digitization can be a time-consuming and costly process, especially for large datasets. One potential solution to this challenge is to use machine learning techniques to automate the digitization process, reducing the need for manual intervention and thereby reducing costs. One area where machine learning can be particularly effective is in enhancing the quality of digital images. In recent years, deep learning-based image super-resolution methods, such as Real-ESRGAN, have shown great promise in improving the quality of low-resolution images [28]. Real-ESRGAN, in particular, is a state-of-the-art deep learning model that can produce high-resolution images with realistic textures and details (Figure 9). Unlike the traditional SRGAN algorithm, it allows for more realistic results closer to reality in an image enhancement project [29]. By using Real-ESRGAN to enhance low-resolution images, organizations can potentially reduce the cost and time required for manual digitization. For example, in the case of historical documents or photographs with low resolution, Real-ESRGAN can be used to create high-quality digital copies that are easier to read and analyze (Figure 7) (Figure 9). This can make the digitization process more efficient, especially when large volumes of data need to be processed.

In addition to reducing costs and improving efficiency, Real-ESRGAN can also have a significant impact on the performance of segmentation models, such as SAM models. Segmentation models rely on accurate and detailed images to accurately identify and classify objects of interest. By enhancing low-resolution images with Real-ESRGAN, segmentation models can be trained on higher-quality data, which can improve their accuracy and reduce false positives.



Figure 9. I) 2022 Istanbul Pan-sharpened Aerial Data Sample, J) 2022 Real-ESRGAN Applied on Pan-sharpened Istanbul Data Sample.

6. Conclusion

In this study, the focus was on increasing the spatial resolution of satellite images through the use of Real-ESRGAN. The results showed that the application of Real-ESRGAN in digitization and segmentation workflows can lead to significant benefits in terms of cost reduction, increased efficiency, and improved accuracy. By using Real-ESRGAN to enhance low-resolution images, organizations can potentially streamline the manual digitization process and reduce costs. Furthermore, the use of Real-ESRGAN can eliminate the need for expensive high-resolution aerial images and reduce reliance on human intervention. However, it is important to note that the quality of the output images produced by Real-ESRGAN can vary based on several factors, and careful consideration and testing are necessary before implementing Real-ESRGAN in a workflow. Overall, Real-ESRGAN represents a powerful tool for machine learning-based image super-resolution techniques, with the potential to transform the way organizations manage and analyze their data.

Acknowledgement

The successful completion of this study would not have been possible without the support and assistance of many individuals and organizations. We would like to express our sincere gratitude to Rukiye Aydın Türkteş, Director of Geographic Information System at İstanbul Metropolitan Municipality, and Hande Çovaç, Deputy Director of Geographic Information System at İstanbul Metropolitan Municipality, for providing valuable data and their continuous support throughout the research process. Their efforts and contributions have been integral to the success of this study. We would also like to extend our appreciation to all the participants who took part in this study and contributed to the research by providing feedback and insights. An abstract of this paper was presented at the TUFUAB XIIth Technical Symposium. The result images applied to all satellite data can be viewed on <https://sehirharitasi.ibb.gov.tr/> in the layers tab under the satellite name 2022.

Funding

This research received no external funding.

Conflicts of interest

The authors declare no conflicts of interest.

References

1. Tsai, R. Y., & Huang, T. S. (1984). Multiframe image restoration and registration. *Advances in Computer Vision and Image Processing*, 1, 317-339.
2. Wang, Y., Fevig, R., & Schultz, R. R. (2008, October). Super-resolution mosaicking of UAV surveillance video. In 2008 15th IEEE International Conference on Image Processing, 345-348. <https://doi.org/10.1109/ICIP.2008.4711762>
3. Zhang, H., Zhang, L., & Shen, H. (2012). A super-resolution reconstruction algorithm for hyperspectral images. *Signal Processing*, 92(9), 2082-2096. <https://doi.org/10.1016/j.sigpro.2012.01.020>
4. Zhang, L., Dong, R., Yuan, S., Li, W., Zheng, J., & Fu, H. (2021). Making low-resolution satellite images reborn: a deep learning approach for super-resolution building extraction. *Remote Sensing*, 13(15), 2872. <https://doi.org/10.3390/rs13152872>
5. Robinson, M. D., Farsiu, S., Lo, J. Y., & Toth, C. A. (2008, October). Efficient restoration and enhancement of super-resolved X-ray images. In 2008 15th IEEE International Conference on Image Processing, 629-632. <https://doi.org/10.1109/ICIP.2008.4711833>
6. Fookes, C., Lin, F., Chandran, V., & Sridharan, S. (2012). Evaluation of image resolution and super-resolution on face recognition performance. *Journal of Visual Communication and Image Representation*, 23(1), 75-93. <https://doi.org/10.1016/j.jvcir.2011.06.004>
7. Ma, L., Zhao, D., & Gao, W. (2012). Learning-based image restoration for compressed images. *Signal processing: Image communication*, 27(1), 54-65. <https://doi.org/10.1016/j.image.2011.05.004>
8. Pickup, L., Roberts, S. J., & Zisserman, A. (2003). A sampled texture prior for image super-resolution. *Advances in neural information processing systems*, 16.
9. Wang, X., Xie, L., Dong, C., & Shan, Y. (2021). Real-EsrGAN: Training real-world blind super-resolution with pure synthetic data. In *Proceedings of the IEEE/CVF international conference on computer vision*, 1905-1914.
10. Richards, J. A. (1984). Thematic mapping from multitemporal image data using the principal components transformation. *Remote Sensing of Environment*, 16(1), 35-46. [https://doi.org/10.1016/0034-4257\(84\)90025-7](https://doi.org/10.1016/0034-4257(84)90025-7)
11. Carper, W., Lillesand, T., & Kiefer, R. (1990). The use of intensity-hue-saturation transformations for merging SPOT panchromatic and multispectral image data. *Photogrammetric Engineering and remote sensing*, 56(4), 459-467.
12. Ranchin, T., & Wald, L. (1993). The wavelet transform for the analysis of remotely sensed images. *International Journal of Remote Sensing*, 14(3), 615-619. <https://doi.org/10.1080/01431169308904362>
13. Zhang, Y. (2004). Understanding image fusion. *Photogrammetric Engineering & Remote Sensing*, 70(6), 657-661.
14. Nasrollahi, K., & Moeslund, T. B. (2014). Super-resolution: a comprehensive survey. *Machine Vision and Applications*, 25, 1423-1468. <https://doi.org/10.1007/s00138-014-0623-4>
15. Peng, Y., Yang, F., Dai, Q., Xu, W., & Vetterli, M. (2012, March). Super-resolution from unregistered aliased images with unknown scalings and shifts. In 2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 857-860. <https://doi.org/10.1109/ICASSP.2012.6288019>
16. Yang, M. C., Huang, D. A., Tsai, C. Y., & Wang, Y. C. F. (2012, July). Self-learning of edge-preserving single image super-resolution via contourlet transform. In 2012 IEEE International Conference on Multimedia and Expo, 574-579. <https://doi.org/10.1109/ICME.2012.169>
17. Yang, Y., & Wang, Z. (2011). A new image super-resolution method in the wavelet domain. In 2011 Sixth International Conference on Image and Graphics, 163-167. <https://doi.org/10.1109/ICIG.2011.79>
18. Liebel, L., & Körner, M. (2016). Single-image super resolution for multispectral remote sensing data using convolutional neural networks. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 41, 883-890. <https://doi.org/10.5194/isprs-archives-XLI-B3-883-2016>
19. Šroubek, F., & Flusser, J. (2006). Resolution enhancement via probabilistic deconvolution of multiple degraded images. *Pattern Recognition Letters*, 27(4), 287-293. <https://doi.org/10.1016/j.patrec.2005.08.010>
20. Ma, W., Pan, Z., Guo, J., & Lei, B. (2018). Super-resolution of remote sensing images based on transferred generative adversarial network. In *IGARSS 2018-2018 IEEE International Geoscience and Remote Sensing Symposium*, 1148-1151. <https://doi.org/10.1109/IGARSS.2018.8517442>
21. Moustafa, M. S., & Sayed, S. A. (2021). Satellite imagery super-resolution using squeeze-and-excitation-based GAN. *International Journal of Aeronautical and Space Sciences*, 22(6), 1481-1492. <https://doi.org/10.1007/s42405-021-00396-6>
22. Wang, J., Gao, K., Zhang, Z., Ni, C., Hu, Z., Chen, D., & Wu, Q. (2021). Multisensor remote sensing imagery super-resolution with conditional GAN. *Journal of Remote Sensing*. <https://doi.org/10.34133/2021/9829706>
23. Zhang, K., Sumbul, G., & Demir, B. (2020, March). An approach to super-resolution of Sentinel-2 images based on generative adversarial networks. In 2020 Mediterranean and Middle-East Geoscience and Remote Sensing Symposium (M2GARSS), 69-72. <https://doi.org/10.1109/M2GARSS47143.2020.9105165>

24. Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., ... & Bengio, Y. (2014). Generative adversarial nets. *Advances in neural information processing systems*, 27.
25. 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*, 4681-4690.
26. Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., ... & Girshick, R. (2023). Segment anything. *Computer Vision and Pattern Recognition*. <https://doi.org/10.48550/arXiv.2304.02643>
27. Wu, Q., & Osco, L. P. (2023). samgeo: A Python package for segmenting geospatial data with the Segment Anything Model (SAM). *Journal of Open Source Software*, 8(89), 5663. <https://doi.org/10.21105/joss.05663>
28. Wang, X., Xie, L., Dong, C., & Shan, Y. (2021). Real-esrgan: Training real-world blind super-resolution with pure synthetic data. In *Proceedings of the IEEE/CVF international conference on computer vision* (pp. 1905-1914).
29. Gazel Bulut, E. B. (2022). Super-Resolution image generation from earth observation satellites using generative adversarial networks. MS Thesis, Hacettepe University.



© Author(s) 2023. This work is distributed under <https://creativecommons.org/licenses/by-sa/4.0/>