Enhancing super-resolution in remote sensing: Integrating GIS data with CNN-based SRGAN models for improved image reconstruction

Ivan Sharshov[1], *Kseniya* Shoshina[1], *Irina* Vasendina[1], *Roman* Aleshko[1], and *Vladimir* Berezovsky[1*]

[1]Northern (Arctic) Federal University named after M.V. Lomonosov, Arkhangelsk, Russia

Abstract. In the field of remote sensing (RS), image super-resolution (SR) techniques play a crucial role across various applications. Traditional SR methods face challenges when applied to long-term coverage datasets with limited spatial resolution. However, recent advancements in deep learning have opened up new possibilities for improving the spatial resolution of RS data. While many convolutional neural network (CNN)-based approaches have achieved excellent performance in developing efficient end-to-end SR models for natural images, they have been less frequently applied to satellite image upscaling with high scale factors. This paper introduces a novel CNN block that enhances the performance of SRGAN-based models. Experimental results show that these architectures benefit from additional data, especially when low-resolution images provide insufficient feature information.

1 Introduction

Remote sensing images have long been utilized in Earth exploration, with various methods developed to recover lost data due to the hardware limitations of satellites and data transmission channels. Recent advancements have significantly improved the resolution and quality of gradient maps, enhancing the accuracy of object classification. With the evolution of satellite and aerial photography, researchers now have access to high-quality surface imagery, which can be used to train and test machine learning algorithms to enhance image quality. Additionally, this has enabled the application of super-resolution algorithms to recover fine details lost during compression.

Super-resolution (SR) methods utilize deep neural networks to train artificial intelligence models that predict objects and elements in low-resolution (LR) images, producing high-resolution (HR) outputs with greater pixel density. Initially, these methods were based on a three-layer autoencoder, as introduced by Dong et al. [1].

In 2015, A. Radford and M. Metz introduced a new architecture for image processing called DCGAN (Deep Convolutional Generative Adversarial Network) [2], which is based on the generative adversarial networks (GAN) framework. While this work was not specifically designed for image upsampling, it had a significant impact on the development

* Corresponding author: v.berezovsky@narfu.ru

of super-resolution techniques. By utilizing the GAN framework, the training process was simplified, reducing the size of the required training dataset.

One of the earliest successful SR architectures using the GAN approach was SRGAN, proposed in 2017 by Ledig et al. [3]. This neural network demonstrated strong performance in restoring details in general-purpose images. However, it had several drawbacks: during training, the network tended to generate artifacts at sharp gradient transitions, and the number and frequency of these artifacts were proportional to the number of RRDB blocks. This limitation prevented the architecture from being used to restore satellite images, as deeper networks are required to reconstruct surface details effectively.

Recent research has focused on minimizing the generation of artifacts in high-resolution images, particularly to improve SRGAN's performance. Wang K. [4] and Rakotonirina N. [5] teams enhanced the architecture, making it more stable during training. The current ESRGAN+ model retains a similar high-level structure but omits BN layers and includes Gaussian noise and RS layers. These modifications have enabled the use of ESRGAN+ as a general-purpose super-resolution algorithm, capable of improving missing parts while preserving the original image's high reconstruction accuracy.

In addition, Jing et al. [6] proposed a new architecture, PEGAN, for super-resolution with high-level image amplification. PEGAN builds on previous ESRGAN work but employs an ensemble of small networks to process the image. It extracts low-frequency components first, which are then processed by two networks: one that extracts structural information via an autoencoder, and another that restores high-frequency features through a multi-pass straight network. This approach achieved near-original image quality when scaling by a factor of 4, but it was insufficient for satellite image processing with a precision of 1 pixel per 8 meters, as the resulting resolution remained less than 1 meter per pixel after amplification.

In 2022, Gao et al. [7] introduced a new method to enhance the spectral resolution of multispectral images by jointly learning a low-rank dictionary pair from overlapping hyperspectral and multispectral regions. However, as the authors noted, their technique is not suitable for large-scale or complex cases.

In light of these challenges, we have proposed a new structure block that utilizes geographic information systems (GIS) data for feature extraction. This block aids SRGAN-based models in improving their performance in single image super-resolution (SISR) tasks, such as upscaling satellite images.

2 Proposed Methods

2.1 Problem statement

The study aimed to explore the correlation between measurement data from the Earth's surface and the original high-resolution (HR) surface images. The primary objective was to enhance the quality and pixel-wise similarity of the reconstructed images using a deep learning algorithm.

One of the main challenges in processing satellite images is the low resolution in the RGB spectrum. As a result, such data cannot be effectively used for image acquisition tasks that are sensitive to image scale, such as image classification and object segmentation on the Earth's surface.

However, numerous satellite monitoring initiatives have recently been launched, collecting data on factors such as reflectance, temperature, segmented water bodies, terrain height, and others. Many of these variables are directly influenced by the objects on the

Earth's surface, leading to the hypothesis that incorporating this data as input in super-resolution algorithms could enhance the reconstruction of low-resolution (LR) images.

2.2 Proposed generator network architecture

To test the hypothesis, we needed to compare two nearly identical neural networks. For this purpose, we employed the original ESRGAN architecture [4], which has demonstrated strong performance in super-resolution (SR) tasks and features a simple pipeline that can be easily modified as required.

We retained the high-level design of ESRGAN (as shown in Figure 1) but introduced a new block for processing collateral data features. This block regulates the parameter of residual scaling within the RRDB blocks for each tensor element. This approach is similar to the attention mechanism proposed by N. Darapeni et al. [8].

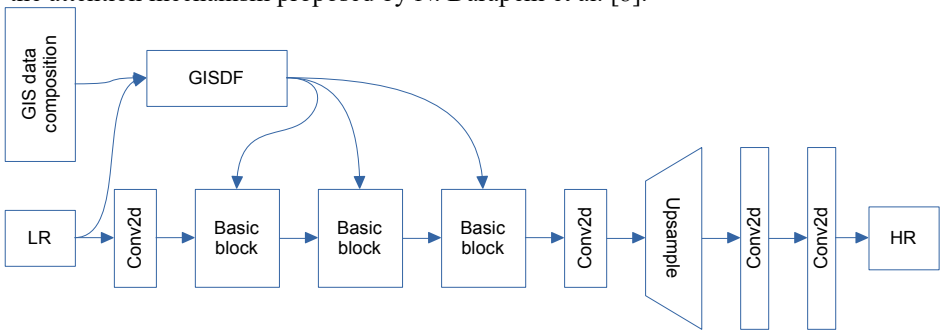

Fig.1. High-level structure of the proposed generator network. We used the basic architecture of ESRGAN and added injection to RRDB block from the GIS data feature extraction block.

The GIS Data Features block (GISDF) extracts features from the low-resolution (LR) composite, which includes the LR image, AVHRR, heat maps, and other layers connected via the channel dimension. This composite is passed through an asymmetric subnetwork, as depicted in Figure 2. The subnetwork extracts features using various convolution kernel sizes and applies the hardtanh activation function to their output, which is then merged via the channel dimension. The result from the GISDF is fed into adapters in every basic block.

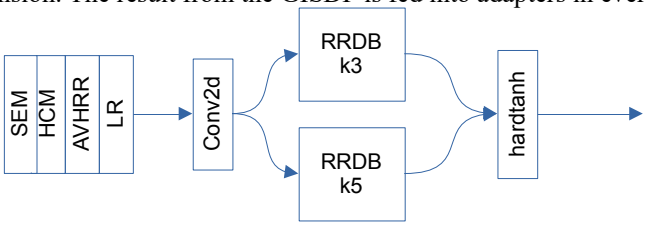

Fig. 2. The structure of GIS data features block.

The adapter is a component of the basic block (originally the RRDB block) and consists of a small feed-forward neural network that also uses the hardtanh activation function (see Figure 3). The output of the adapter is used as residual scaling, replacing the original constant residual scaling. As a result, this block functions as a "gate" for the basic block, allowing pre-processing during the forward pass. If the adapter's tensor output is close to zero, the basic block can be skipped, reducing the generation time of the SR image.

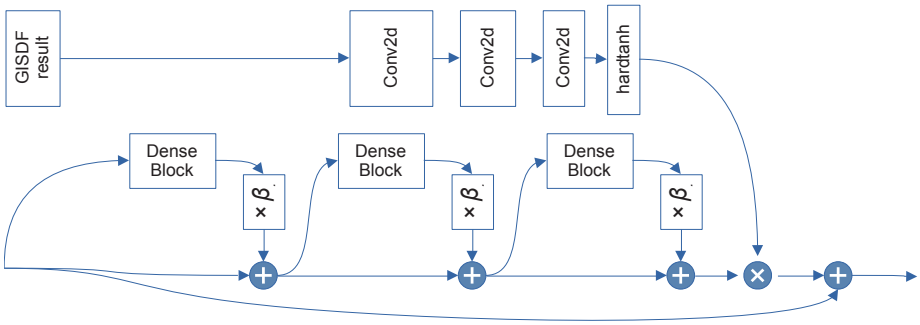

Fig. 3. The adapter and RRDB structure.

3 Datasets and experimental setup

3.1 Datasets

The high-resolution (HR) images used for the single image super-resolution (SISR) experiments were sourced from the USGS collection of high-resolution orthorectified images [9]. These images combine the visual characteristics of aerial photographs with the geometric precision of a map. An orthoimage is a uniformly scaled image that has been corrected for distortions such as building tilt and scale variations caused by terrain, sensor geometry, and camera angle. The dataset includes black and white, natural color, and color infrared images with a resolution of 1 meter or smaller per pixel. For the experiments, we selected various locations across the United States, aiming to match diverse surface types in natural color. The best surface matches were found in Alaska and the eastern regions, as depicted in Figure 4.

Fig. 4. Overview of high-resolution images dataset

We utilized several map databases as sources of collateral data (Figure 5). The first dataset comes from normalized difference vegetation index (NDVI) composites, which are produced from multiple daily observations by the advanced very high-resolution radiometer (AVHRR). These observations are composited to create near cloud-free images that display maximum greenness [10]. An NDVI ratio is generated from bands one and two of the AVHRR composite, resulting in a five-band dataset with a spatial resolution of 1 km.

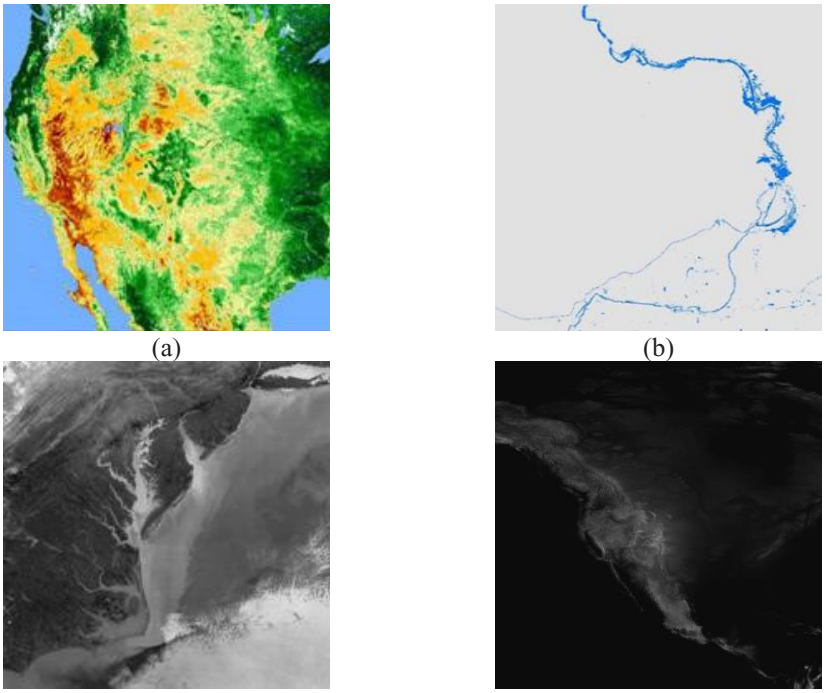

Fig. 5. Overview of four datasets of collateral data used in the research: **(a)** VHRR **(b)** DSWE **(c)** HCMM **(d)** GMTED2010

The second collateral dataset is the Landsat Collection 2 Level-3 Dynamic Surface Water Extent (DSWE) [11], which includes six raster data products related to surface water existence and condition. The Landsat C2 DSWE products are processed to a 30-meter spatial resolution using the Albers Equal Area (AEA) projection and the World Geodetic System 1984 (WGS84) datum, gridded to a common tiling scheme.

The third dataset was sourced from the Heat Capacity Mapping Mission (HCMM) [12], part of NASA's Applications Explorer Missions (AEM-A). The mission provided day/night coverage of specific areas at intervals of 12 to 36 hours, with a 16-day repeat cycle. The HCMM sensor operated in two channels: one for visible to near-infrared radiation (0.55–1.1 micrometers) and another for thermal infrared radiation (10.5–12.5 micrometers). The HCMM dataset includes day visible, day thermal infrared, night infrared, temperature difference, and apparent thermal inertia images. The latter two are approximations derived from radiometer measurements.

Additionally, we used digital elevation data from the Global Multi-resolution Terrain Elevation Data 2010 (GMTED2010) [13]. This dataset provides surface height data with a spatial resolution of 7.5 arc-seconds and 1-meter vertical resolution. The global vertical accuracy of GMTED2010 is measured by the Root Mean Square Error (RMSE), with an estimated global accuracy of 6 meters RMSE, according to data from the National Geospatial-Intelligence Agency.

All data sources were aligned with the HR image layers, which were divided into 512-meter chunks with a spatial resolution of 1 meter per pixel. The collateral data were cropped to the same 512-meter blocks and scaled to 128 pixels. The data were then normalized to a range of zero to one and saved in float32 format. The corresponding low-resolution (LR) images were generated in real-time using bicubic interpolation.

3.2 Experimental setup

In two separate experiments, datasets were collected to perform distinct tasks. The first experiment focused on evaluating the raw performance of a 24-block ESRGAN model at a scale factor of x4, with a patch size of 128 pixels. This choice was based on the original work [4], which identified x4 scaling with patch sizes of 96, 128, and 192 pixels as effective starting points and provided statistical data for comparison. The second experiment sought to explore correlations between collateral data and high-resolution (HR) images, with the aim of improving generator performance.

The training process for each neural network consisted of two stages. First, both the generator and discriminator underwent a pretraining phase using the L1 loss function:

$$Loss_{L1} = \frac{1}{n}\sum_{i=1}^{n}\left\|F_i^{SR} - F_i^{HP}\right\| \quad (1)$$

Next, adversarial learning was implemented according to the methodology proposed by Goodfellow et al. [14].

For optimization, we utilized the Adam optimizer [15] with an initial learning rate of 10^{-4}. Each model was trained for 100 epochs, with the learning rate reduced by a factor of 0.9 every four epochs throughout the training process. The models were trained using PyTorch on an RX7900XTX GPU with 24 GB of dedicated memory.

4 Results

The evaluation results are presented in Table I. GESRGAN demonstrated the ability to generate finer details, as seen in Figure 6 for the "3919651 ortho611 06 — APPA SBR 2009 200910 sbr appalachian trail appa ga-nc-tn-va 0x3000m utm17 cnir 611 ortho611 06" crop.

Our findings suggest that processing collateral data through RRDB filters and using it as a residual scaling factor for basic blocks in the SRGAN framework can enhance the performance of GAN-based networks for super-resolution image reconstruction. However, further research is needed to investigate the impact of different types of collateral data and their processing methods.

Improvements could be achieved by increasing the patch size while maintaining the same number of basic blocks. Enlarging the size of the low-resolution (LR) image may improve learning stability and the peak signal-to-noise ratio (PSNR) of the generated super-resolution (SR) image, as suggested in previous work [4]. Future experiments could utilize 256 to 512 pixel patches to gather additional statistics and conduct comparative studies.

Table 1. Comparison of PSNR and SSIM of super-resolution models trained on orthoimage high resolution dataset width x8 scale factor.

Model	ESRGAN	GESRGAN (ours)
PSNR	27.3630	27.3986
SSIM	0.9632	0.9673

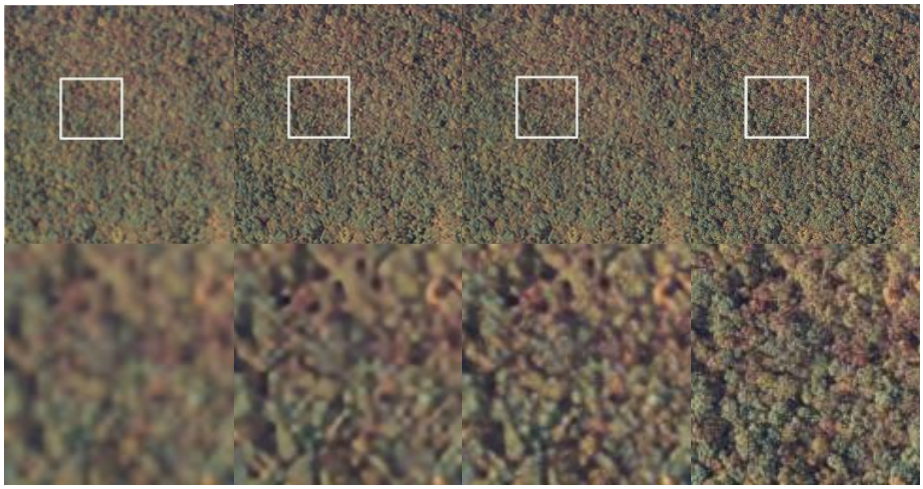

Fig. 6. Visual comparisons between (a) bicubic interpolation, (b) ESRGAN SR image, (c) GESRGAN SR image, (d) original image.

The number of residual connections also plays a significant role, as an increased number of residual links can enhance both performance and complexity in neural networks [16]. This could be achieved by incorporating additional residual-dense blocks within the RRDB structure.

Another way to improve performance is by increasing the complexity of the GIS Data Features block. In the current study, only two RRDB blocks with different convolution kernel sizes were used. Further research should focus on finding a more optimal structure for this block and analyzing the overall impact of each component of the computational sub-graph on the performance of the adapter block and the network as a whole.

As demonstrated by L. Baozhong and C. Ji [17], using an attention mechanism can significantly improve performance in SR tasks. Therefore, we propose modifying the adapter block in our basic blocks to remove residual scaling and instead use feature information from the GIS data block to generate channel and spatial attention masks.

5 Conclution

In conclusion, this study aimed to enhance the performance of deep learning models for super-resolution image reconstruction by investigating the role of various types of collateral data and their processing. The results indicate that processing collateral data through RRDB filters and using it as a residual scaling factor within the SRGAN framework can significantly improve model performance. Further research is necessary to explore the effects of different types of collateral data and processing methods, including increasing patch size, incorporating residual-dense blocks (RDB) within RRDB, and implementing attention mechanisms. Additionally, using larger patches and increasing the complexity of the GIS data features block could further enhance model performance. Overall, the study presents promising advancements in super-resolution image reconstruction, highlighting the need for continued research to refine and improve these models.

The research was supported by the Russian Science Foundation and the Government of the Arkhangelsk Region, project No. 22-11-20025..

References

1. C. Dong, C. C. Loy, K. He, X. Tang, *Image super-resolution using deep convolutional networks*, IEEE Transactions on Pattern Analysis and Machine Intelligence, **38**(2), 295–307 (2016)
2. A. Radford, L. Metz, S. Chintala, Unsupervised representation learning with deep convolutional generative adversarial networks, arXiv preprint arXiv:1511.06434, Available: https://arxiv.org/pdf/1511.06434v1.pdf (2015)
3. C. Ledig *et al.*, *Photo-realistic single image super-resolution using a generative adversarial network*, in 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 105–114 (2017)
4. X. Wang *et al.*, *ESRGAN: enhanced super-resolution generative adversarial networks*, LNCS **11133** 63–79 (2019)
5. N. C. Rakotonirina, A. Rasoanaivo, *ESRGAN+ : further improving enhanced super-resolution generative adversarial network*, in ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP*)*, 3637–3641 (2020)
6. C. -W. Jing, Z. -X. Huang, Z. -Y. Ling, *An image super-resolution reconstruction method based on PEGAN*, IEEE Access **11**, 102550-102561, (2022)
7. L. Gao, D. Hong, J. Yao, B. Zhang, P. Gamba, J. Chanussot, *Spectral superresolution of multispectral imagery with joint sparse and low-rank learning*, IEEE Trans. Geosci. Remote Sens. **59**(3), 2269–2280 (2021)
8. N. Darapaneni *et al.*, *Contextual attention mechanism, SRGAN based inpainting system for eliminating interruptions from images,* c2204.02591 Available: https://arxiv.org/pdf/2204.02591.pdf (2022)
9. *USGS EROS archive - aerial photography - high resolution orthoimagery (HRO),* doi: 10.5066/F73X84W6 (2018)
10. *USGS EROS Archive - AVHRR normalized difference vegetation index (ndvi) composites*, doi: 10.5066/F7707ZKN.(2018)
11. *Landsat collection 2 level-3 dynamic surface water extent (DSWE) contains six acquisition-based raster data products pertaining to the existence and condition of surface water*, doi: 10.5066/P9DPWBUS (2022)
12. *USGS EROS archive - heat capacity mapping mission (HCMM)*, doi: 10.5066/P906HJ0V (2018)
13. *USGS EROS archive - digital elevation - global multi-resolution terrain elevation data 2010 (GMTED2010)*, doi: doi.org/10.5066/F7J38R2N (2018).
14. I. J. Goodfellow *et al.*, *Generative adversarial nets*. arXiv preprint arXiv:1406.2661 Available: https://arxiv.org/pdf/1406.2661.pdf (2014)
15. D. Kingma, J. Ba, *Adam: a method for stochastic optimization*, International Conference on Learning Representations, arXiv preprint arXiv:1412.6980, Available: https://arxiv.org/pdf/1412.6980 (2014)
16. Y. Tai, J. Yang, X. Liu, *Image super-resolution via deep recursive residual network* ,In Proceedings of the IEEE conference on computer vision and pattern recognition, 3147-3155 (2017)
17. B. Liu, J. Chen, *A super resolution algorithm based on attention mechanism and SRGAN network*, IEEE Access **9**, 139138-139145 (2021)