Vol. 24, No. 1, pp. 9-17 (2024)
Journal of Agricultural Physics
ISSN 0973-032X
http://www.agrophysics.in

Research Article

Super Resolution of Sentinel-2 Imagery using Deep Learning Model: SRAttentionNet

BANOTH JAGDISH NAIK[1], ANSHU BHARADWAJ[1]*, VINAY KUMAR SEHGAL[2], MUKESH KUMAR[1], SHASHI DAHIYA[1] AND RAJNI JAIN[6]

[1]*Division of Computer Application, The Graduate School, ICAR-Indian Agricultural Research Institute, New Delhi-110012*
[2]*Division of Agricultural Physics, ICAR-Indian Agricultural Research Institute, New Delhi-110012*
[3]*ICAR-National Institute of Agricultural Economics and Policy Research, New Delhi-110012*

ABSTRACT

Sentinel-2 satellite imagery plays a crucial role in various Earth observation applications, but its spatial resolution is limited. Super-resolution techniques aim to enhance the resolution of Sentinel-2 images, enabling finer details to be observed. This paper presents a novel deep learning model, Super Resolution Attention Network (SRAttentionNet), designed for super-resolution of Sentinel-2 images. The model incorporates attention mechanisms within residual blocks to selectively focus on important features, leading to improved reconstruction quality. This paper evaluates SRAttentionNet on the SEN2VENμS dataset and compares its performance with state-of-the-art models like Deep Sentinel-2(DSen2), Enhanced Deep Super-Resolution (EDSR), and Efficient Sub-Pixel Convolutional Neural Network(ESPCN). The results demonstrate that SRAttentionNet outperforms these models in terms of Peak Signal to Noise ratio(PSNR), Structural Similarity(SSIM), and visual quality, highlighting the effectiveness of attention mechanisms in super-resolution tasks.

Key words: Sentinel-2, Super-resolution, Attention mechanism, Peak signal to noise ratio, Structural similarity

Introduction

Sentinel-2 is a constellation of two Earth observation satellites operated by the European Space Agency (ESA) as part of the Copernicus Programme. These satellites provide high-resolution multispectral imagery of the Earth's surface, covering a wide range of applications such as land monitoring, agriculture, forestry, and disaster management. However, the spatial resolution of Sentinel-2 images varies across different spectral bands, with some bands having lower resolution than others. This limitation can hinder the ability to discern fine details in the imagery.

*Corresponding author,
Email: Anshu.Bharadwaj@icar.gov.in

Super-resolution techniques offer a solution to this problem by enhancing the resolution of lower-resolution bands to match that of the highest resolution bands. The evolution of super-resolution for satellite images has progressed significantly, transitioning from traditional interpolation methods to sophisticated deep learning models. Initially, basic techniques like bicubic interpolation were used, but they often produced blurry results with limited detail enhancement.

The advent of deep learning revolutionized the field, with models like Super-Resolution Convolutional Neural Network (SRCNN) pioneering the use of Convolutional Neural Networks (CNNs) (Dong *et al.*, 2015, 2016) for super-resolution. This

marked a shift towards learning-based approaches that could capture complex patterns and relationships in image data. Subsequent models like Fast Super-Resolution Convolutional Neural Network (FSRCNN) (Passarella *et al*., 2022) and Very Deep Super-Resolution (VDSR) (Kim *et al*., 2016) further refined CNN architectures, focusing on efficiency and depth to improve performance.

The introduction of generative adversarial networks (GANs) with SRGAN [Nagano and Kikuta, 2018] brought a new dimension to super-resolution, enabling the generation of visually realistic high-resolution images with finer details and textures. Attention mechanisms, as seen in Residual Channel attention Networks (RCAN) (Zhang *et al.*, 2018a) and second order Attention Network (SAN) (Dai *et al.*, 2019), further enhanced the ability of models to selectively focus on important features, leading to even better reconstruction quality.

Recent advancements have explored various architectural innovations, such as Residual Dense Networks (RDN) (Zhang *et al.*, 2018b), Cascading Residual Networks (CARN) (Ahn *et al*., 2018), and enhanced GANs (ESRGAN) (Lin, 2022), to push the boundaries of super-resolution performance. Additionally, techniques like Meta-learning Super Resolution (Meta-SR) (Hu *et al*., 2019) and progressive super-resolution (ProSR) (Hajian and Aramvith, 2023; Kim *et al*., 2019) have been introduced to address challenges like adaptability and stability.

The evolution of super-resolution for satellite images continues to be an active research area, with ongoing efforts to develop more sophisticated models that can produce high-quality, detailed images for various applications in Earth observation and remote sensing.

In this paper, a novel deep learning model called SRAttentionNet for super-resolution of Sentinel-2 images has been proposed. This model incorporates attention mechanisms within residual blocks, allowing it to selectively focus on important features during the reconstruction process. This attention mechanism enables the model to better preserve fine details and textures, leading to improved image quality.

SRAttentionNet is evaluated on the SEN2VENμS dataset, which consists of paired Sentinel-2 and VENμS satellite images. VENμS images have a higher spatial resolution than Sentinel-2 images, making them suitable as reference images for super-resolution. The performance of SRAttentionNet is compared with state-of-the-art models like Bicubic (Liu *et al*., 2013), DSEN2 (Lanaras *et al*., 2018), EDSR (Lin, 2022), and ESPCN (Shi *et al*., 2016). The results demonstrate that SRAttentionNet outperforms these models in terms of peak signal-to-noise ratio (PSNR) (Hore and Ziou, 2010), structural similarity index measure (SSIM) (Hore and Ziou, 2010), and visual quality.

Related Work

Super-resolution of satellite imagery has been an active research area in recent years. Dong *et al.* (2015 & 2016) introduced the concept of using deep convolutional neural networks for super-resolution, demonstrating their effectiveness in learning the mapping between low-resolution and high-resolution images. Kim *et al.* (2016) explored the use of very deep networks for super-resolution, showing that increasing the network depth can lead to better performance. Ledig *et al.* (2017) introduced the use of generative adversarial networks (GANs) for super-resolution, producing visually appealing high-resolution images with finer details and textures. Li *et al*. (2019) introduced super-resolution feedback network (SRFBN) to refine low-level representations with high-level information. Zhang *et al.* (2018a,b) proposed this model, incorporating channel attention mechanisms into a residual network architecture, enabling it to selectively focus on important features and achieve state-of-the-art performance. Dai *et al.* (2019) introduced this model, extending the attention mechanism to second-order interactions, further improving the performance of super-resolution networks. Wang *et al.* (2018, 2021) improved upon SRGAN by incorporating several enhancements, such as a relativistic discriminator and a perceptual loss function, resulting in even better visual quality. Tong *et al.* (2017) introduced a novel single image super resolution method by introducing dense skip connections to boost reconstruction performance.

Material and Methods

Proposed Method

The proposed model, SRAttentionNet, is a deep learning model designed for super-resolution of Sentinel-2 images. The model architecture is based on a residual network (ResNet), which consists of multiple residual blocks. Each residual block contains two convolutional layers with a skip connection. This skip connection allows the network to learn residual functions, making it easier to train deeper networks.

Attention mechanism is incorporated within each residual block to selectively focus on important features. The attention mechanism consists of two parts: a channel attention module and a spatial attention module. The channel attention module learns to weight the importance of different channels, while the spatial attention module learns to weight the importance of different spatial locations.

The overall architecture of SRAttentionNet is shown in Fig. 2. The model takes a low-resolution Sentinel-2 image as input and produces a high-resolution image as output. The input image is first passed through a convolutional layer to extract initial features. These features are then passed through multiple residual blocks with attention. Finally, the features are upsampled using a pixel shuffle layer to produce the high-resolution output image.

Dataset

SEN2VENμS (Michel *et al*., 2022) is an open dataset (available at https://zenodo.org/records/6514159) designed for the super-resolution of Sentinel-2 images by leveraging simultaneous acquisitions with the VENμS satellite. The dataset consists of 10m and 20m cloud-free surface reflectance patches from Sentinel-2, with their reference spatially-registered surface reflectance patches at 5 meters resolution acquired on the same day by the VENμS satellite. The dataset covers 29 locations with a total of 132,955 patches of 256×256 pixels at 5m resolution and is intended for the training of super-resolution algorithms to enhance the spatial resolution of the Sentinel-2 bands to 5m. Preparation of training data for super-resolution tasks using the Sen2Venus dataset includes the input and target tensors.

Input tensor: Low-resolution Sentinel-2 data (10m or 20m bands, or a combination of both).

Target tensor: High-resolution Venus data (5m bands). The goal of the super-resolution task is to learn a model that can take the low-resolution Sentinel-2 data as input and produce an output that matches the high-resolution Venus data. During training, the model will be optimized to minimize the difference between its output and the target tensor (high-resolution Venus data) using a suitable loss function, such as Root Mean Squared Error (RMSE) , and the dataset files are stored as PyTorch tensor files with the *.pt* extension. Fig. 1 shows different bands of input images and target images.

The image patches are encoded as ready-to-use tensors serialized by the PyTorch library. The tensors have the shape [n, c, w, h] where n is the number of patches, c=4 is the number of spectral bands, w is the patch width and h is the patch height. The tensor values are encoded as 16-bit signed integers to save storage space. To convert them back to floating point surface reflectance values, each value needs to be divided by 10,000 after loading the tensor. There are separate tensor files for the different spatial resolutions (5m, 10m, 20m) and spectral band groups of the Sentinel-2 data. The naming convention for the tensor files follows the pattern {id}_{resolution}_{bands}.pt, e.g. ALSACE_32TLR_20190822_05m_b2b3b4b8.pt contains 5m resolution patches for Sentinel-2 bands 2, 3, 4, 8.

Network Architecture

SRAttentionNet, the proposed model for super-resolution of multi-spectral imagery, features a comprehensive architecture optimized for high-performance and efficiency. Fig. 2 represents the architecture of the proposed model. It begins with an initial convolution layer, followed by 32 Residual Blocks with Attention, a mid-convolution layer with a global residual connection, an upscaling module using PixelShuffle, and a final convolution layer. Key features of SRAttentionNet include:

a. **Residual Blocks with Attention**: Each block contains two convolutional layers, incorporates a channel-wise attention mechanism, uses Leaky ReLU activation with a slope of 0.2, and applies a scaling factor of 0.1 to the output.

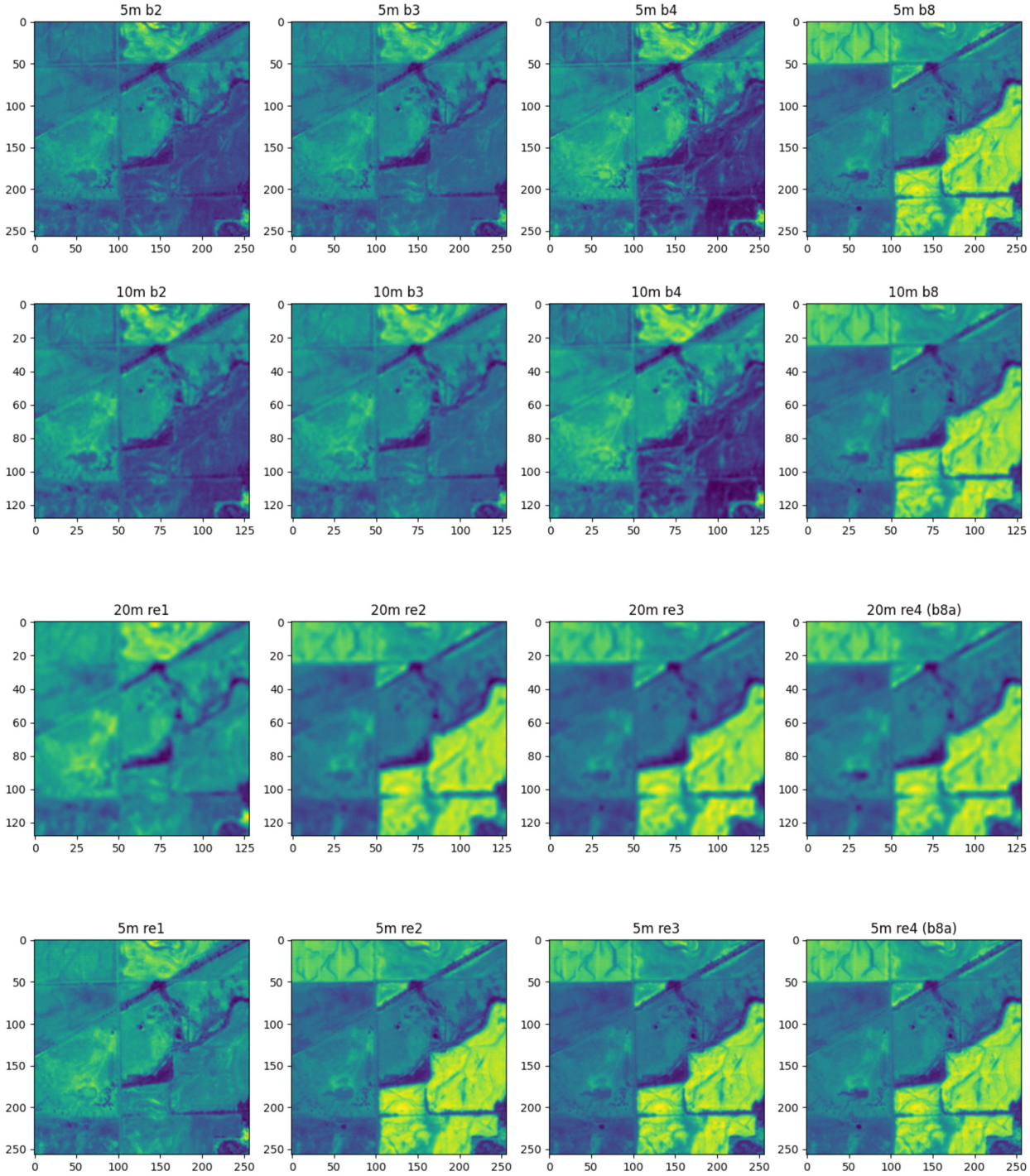

Fig. 1. First row-5m band Sentinel-2, Second row-10m bands Sentinel-2, third-20m bands Sentinel-2 and last row 5m bands VENμS (from top to bottom)

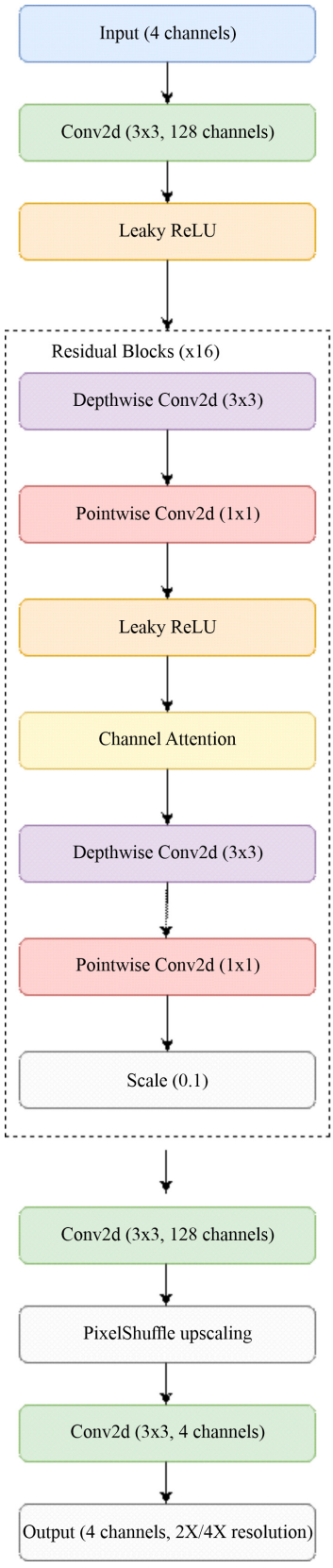

Fig. 2. Architecture of SRAttentionNet

b. **Dense Structure**: The model utilizes 32 residual blocks, and 256 feature channels, allowing for richer feature learning.

c. **Attention Mechanism**: The channel-wise attention in each residual block helps the network focus on the most important features, enhancing the detail preservation and artifact suppression.

d. **Upscaling**: The PixelShuffle technique is employed for efficient and effective upscaling, minimizing artifacts compared to traditional methods.

e. **Input/Output**: Designed to handle 4-channel input/output, SRAttentionNet is tailored for multi-spectral imagery, such as those from Sentinel-2 satellites.

Evaluation Metrics

SSIM (Structural Similarity Index Measure)

SSIM is used to measure the similarity between two images. It is designed to improve upon traditional metrics like PSNR and MSE by considering changes in structural information (Captures the structures within an image, such as edges and textures). Higher the value better performance.

It is calculated using the following formula, where x and y are the two images being compared:

$$\text{SSIM}(x, y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)}$$

- μ_x and μ_y: Mean intensities of images x and y
- s_x^2 and s_μ^2: Variances of images and
- s_{xy}: Covariance of x and y
- c_1 and c_2: Small constants to avoid division by zero.
- SSIM values ranges from -1 to 1, where 1 indicates perfect similarity

PSNR (Peak Signal-to-Noise Ratio)

PSNR measures the ratio between the maximum possible power of a signal (image) and the power of corrupting noise that affects the fidelity of its representation. It is commonly used to quantify the quality of reconstructed images. Higher the value better performance.

$$\text{PSNR} = 10 \log_{10} \left(\frac{\text{MAX}_I^2}{\text{MSE}} \right)$$

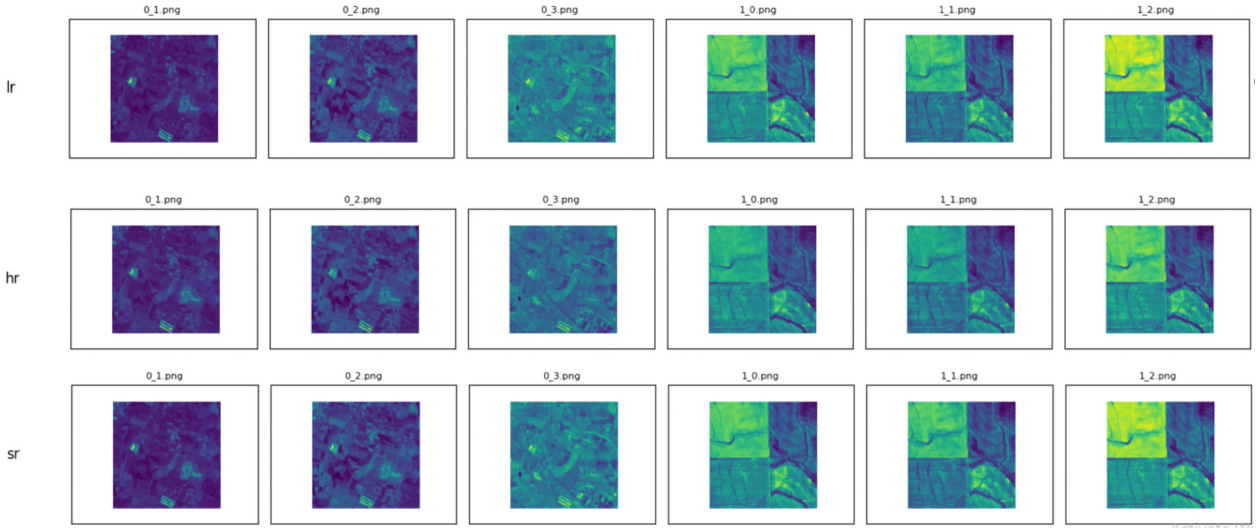

Fig. 3. Model predictions for 2x resolution. First row shows input low resolution images, second row shows high resolution target images and third row shows the predicted high resolution images (from top to bottom)

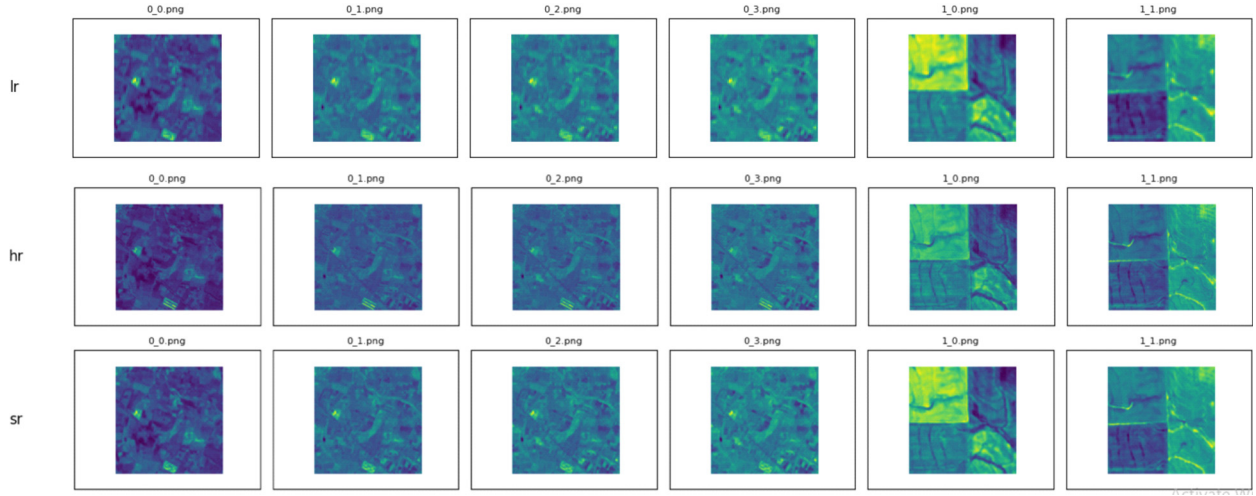

Fig. 4. Model predictions for 4x resolution. First row shows input low resolution images, second row shows high resolution target images and third row shows the predicted high resolution images (from top to bottom)

- MAXI: Maximum possible pixel value of the image (e.g., 255 for 8-bit images).

Root Mean Squared Error (RMSE)

Root Mean Squared Error between the original and reconstructed image, calculated as:

$$RMSE = \sqrt{\frac{1}{mn}\sum_{i=0}^{m-1}\sum_{j=0}^{n-1}[\mathrm{I}(i,j) - K(i,j)]^2}$$

- I and K: Original and reconstructed images
- m and n: Dimensions of the images.

Normalized Cross Correlation Coefficient (NCC)

This metric is useful for the quantification of image similarity. Higher the value better performance.

$$\mathrm{NCC}(x,y) = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_i (x_i - \bar{x})^2 \sum_i (y_i - \bar{y})^2}}$$

- x_i and y_i : Pixel values of the images x and y
- $\bar{x}$ and $\bar{y}$: Mean pixel values of x and y
- NCC values range from -1 to 1, where 1 indicates perfect correlation.

Error of Relative Global Adimensional Synthesis (ERGAS)

It is a quality evaluation method proposed for image fusion research, which reflects the degree of spectral distortion between the restored image and the reference image. Lower the value better performance.

$$\text{ERGAS} = 100\frac{h}{l}\sqrt{\frac{1}{K}\sum_{k=1}^{K}\left(\frac{RMSE(k)}{\mu(k)}\right)^2}$$

- l and h represent the resolution before and after image reconstruction
- k represents the number of bands
- $\mu(k)$ represents the average of band
- *RMSE* represents the root mean square error of the image.

Spectral Angle Mapper (SAM)

Spectral Angle Mapper [Chakravarty *et al*., 2021] determines the spectral similarity between image spectra and reference spectra. Lower the value better performance.

SSIM Loss

It is used in combination with NCC to train the model, aiming to better preserve structural information in the reconstructed images.

SSIM Loss = 1 – SSIM (x, y)

This transforms the SSIM index into a loss function where a lower value indicates better performance.

$$SAM = \frac{1}{HW}\sum_{i=0}^{H}\sum_{j=0}^{\mathrm{W}}\theta(\mathrm{X}(i,j),Y(i,j))$$

Metrics like PSNR, SSIM, NCC, RMSE, SAM, ERGAS were used as quantitative and qualitative metrics to evaluate the quality of the reconstructed images.

Training Details

All models were trained on an NVIDIA RTX 3090 GPU. Each model underwent training for 10 epochs to ensure a fair comparison. We train SRAttentionNet using the Adam optimizer with a learning rate of 0.0001. A batch size of 32 and the L1 loss function is used to optimize the model.

Results and Discussion

The quantitative results of our experiments are shown in Table 1 for 2x and Table 2 for 4x. SRAttentionNet outperforms all other models in terms of PSNR and SSIM. This demonstrates the effectiveness of attention mechanisms in improving the quality of super-resolved Sentinel-2 images.

The visual comparisons of the reconstructed images are also provided in Figure 2. SRAttentionNet produces sharper and more detailed images than the other models. This is particularly evident in areas with fine textures, such as vegetation and urban areas.

Table 1. Comparision results for 2x Resolution

Model	PSNR	SSIM	NCC	RMSE	SAM	ERGAS
Bicubic	45.5413	0.9882	0.9463	0.0056	0.6546	0.6647
DSen2	46.2345	0.9910	0.9576	0.0051	0.5904	0.5979
EDSR	46.7697	0.9915	0.9594	0.0048	0.5695	0.5746
ESPCN	45.9959	0.9902	0.9528	0.0053	0.6206	0.6261
SRAttentionNet	47.4430	0.9929	0.9661	0.0044	0.5305	0.5354

Table 2. Comparision results for 4x Resolution

Model	PSNR	SSIM	NCC	RMSE	SAM	ERGAS
Bicubic	41.9953	0.9669	0.9111	0.0083	0.6745	0.7623
DSen2	43.3323	0.9718	0.9305	0.0071	0.5719	0.6546
EDSR	43.4606	0.9754	0.9236	0.0070	0.5660	0.6475
ESPCN	42.6679	0.9713	0.9218	0.0077	0.6132	0.7087
SRAttentionNet	44.3811	0.9802	0.9446	0.0063	0.5143	0.5821

Conclusion

In this paper, a novel deep learning model-SRAttentionNet, for super-resolution of Sentinel-2 images is proposed. This model incorporates attention mechanisms within residual blocks to selectively focus on important features, leading to improved reconstruction quality. SRAttentionNet is evaluated on the SEN2VENμS dataset and compared its performance with state-of-the-art models like Bicubic, DSEN2, EDSR, and ESPCN. The results demonstrate that SRAttentionNet outperforms these models in terms of PSNR, SSIM, and visual quality, highlighting the effectiveness of attention mechanisms in super-resolution tasks.

References

Ahn, N., Kang, B. and Sohn, K.A. 2018. Fast, accurate, and lightweight super-resolution with cascading residual network. P. 252-268. In *Proceedings of the European conference on computer vision (ECCV)*.

Chakravarty, S., Paikaray, B.K., Mishra, R. and Dash, S. 2021. Hyperspectral image classification using spectral angle mapper. P. 87-90. In *2021 IEEE International Women in Engineering (WIE) Conference on Electrical and Computer Engineering (WIECON-ECE)*. *IEEE*.

Dai, T., Cai, J., Zhang, Y., Xia, S.T. and Zhang, L. 2019. Second-order attention network for single image super-resolution. P. 11065-11074. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*.

Dong, C., Loy, C.C. and Tang, X. 2016. Accelerating the super-resolution convolutional neural network. P. 391-407. In *Computer Vision-ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part II 14*. *Springer International Publishing*.

Dong, C., Loy, C.C., He, K. and Tang, X. 2015. Image super-resolution using deep convolutional networks. *IEEE PAMI* **38**: 295-307.

Hajian, A. and Aramvith, S. 2023. Fusion objective function on progressive super-resolution network. *Journal of Sensor and Actuator Networks* **12**: 26.

Hore, A. and Ziou, D. 2010, August. Image quality metrics: PSNR vs. SSIM. P. 2366-2369. In *2010 20th international conference on pattern recognition*. *IEEE*.

Hu, X., Mu, H., Zhang, X., Wang, Z., Tan, T. and Sun, J. 2019. Meta-SR: A magnification-arbitrary network for super-resolution. P. 1575-1584. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*.

Kim, D., Kim, M., Kwon, G. and Kim, D.S. 2019. Progressive face super-resolution via attention to facial landmark. *arXiv preprint arXiv:1908.08239*.

Kim, J., Lee, J.K. and Lee, K.M. 2016. Accurate image super-resolution using very deep convolutional networks. P. 1646-1654. In *Proceedings of the IEEE conference on computer vision and pattern recognition*.

Lanaras, C., Bioucas-Dias, J., Galliani, S., Baltsavias, E. and Schindler, K. 2018. Super-resolution of Sentinel-2 images: Learning a globally applicable deep neural network. ISPRS J. Photogramm. *Remote Sensing* **146**: 305-319.

Ledig, C., Theis, L., Huszár, F., Caballero, J., Cunningham, A., Acosta, Aitken, A., Tejani, A., Tetz, J., Wang, Z. and Shi, W. 2017. Photo-realistic single image super-resolution using a generative adversarial network. P. 4681-4690. In *Proceedings of the IEEE conference on computer vision and pattern recognition*.

Li, Z., Yang, J., Liu, Z., Yang, X., Jeon, G. and Wu, W. 2019. Feedback network for image super-resolution. P. 3867-3876. In *Proceedings of the*

IEEE/CVF conference on computer vision and pattern recognition.

Lin, K. 2022. The performance of single-image super-resolution algorithm: EDSR. P. 964-968. In *2022 IEEE 5th international conference on information systems and computer aided education (ICISCAE)*. *IEEE*.

Liu, J., Gan, Z. and Zhu, X. 2013. Directional bicubic interpolation—A new method of image super-resolution. P. 463-470. In *3rd International Conference on Multimedia Technology (ICMT-13)*. *Atlantis Press*.

Michel, J., Vinasco-Salinas, J., Inglada, J. and Hagolle, O. 2022. Sen2venμs, a dataset for the training of sentinel-2 super-resolution algorithms. *Data* **7**: 96.

Nagano, Y. and Kikuta, Y. 2018. SRGAN for super-resolving low-resolution food images. P. 33-37. In *Proceedings of the Joint Workshop on Multimedia for Cooking and Eating Activities and Multimedia Assisted Dietary Management*.

Passarella, L.S., Mahajan, S., Pal, A. and Norman, M.R. 2022. Reconstructing high resolution ESM data through a novel fast super resolution convolutional neural network (FSRCNN). *Geophysical Research Letters* **49**(4).

Shi, W., Caballero, J., Huszár, F., Totz, J., Aitken, A. P., Bishop, R., Rueckert, D. and Wang, Z. 2016. Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network. P. 1874-1883. In *Proceedings of the IEEE conference on computer vision and pattern recognition*.

Tong, T., Li, G., Liu, X. and Gao, Q. 2017. Image super-resolution using dense skip connections. P. 4799-4807. In *Proceedings of the IEEE international conference on computer vision*.

Wang, L., Shen, J., Tang, E., Zheng, S. and Xu, L. 2021. Multi-scale attention network for image super-resolution. *Journal of Visual Communication and Image Representation* **80**: 103300

Wang, X., Yu, K., Wu, S., Gu, J., Liu, Y., Dong, C., Qiao, Y. and Change Loy, C. 2018. Esrgan: Enhanced super-resolution generative adversarial networks. In *Proceedings of the European conference on computer vision (ECCV) workshops*.

Zhang, Y., Li, K., Li, K., Wang, L., Zhong, B. and Fu, Y. 2018. Image super-resolution using very deep residual channel attention networks. P. 286-301. In *Proceedings of the European conference on computer vision (ECCV)*.

Zhang, Y., Tian, Y., Kong, Y., Zhong, B. and Fu, Y. 2018. Residual dense network for image super-resolution. P. 2472-2481. In *Proceedings of the IEEE conference on computer vision and pattern recognition*.

Received: 5 April 2024; Accepted: 18 June 2024