Super Resolution of Resourcesat-2 Satellite Imagery using SRCNN

Mr. A. Nikhil Karthik[1]
1. B.Tech. Student, Dept. of CSE, Keshav Memorial Institute of Technology, India
agastyakarthik29@gmail.com

1. Abstract

Remote sensing technologies play a pivotal role in monitoring and understanding our changing environment. However, acquiring high-resolution satellite imagery poses challenges, often limited by cost and availability. This research addresses the critical need for enhanced spatial resolution in satellite imagery through the application of Convolutional Neural Networks (CNNs)[1]. Specifically, we focus on the super-resolution of Resourcesat-2 LISS-3 imagery, a widely utilized remote sensing dataset. Given the 24-meter resolution of the original imagery, obtaining higher-resolution (5-meter) data is a daunting task. In response to this challenge, our proposed method leverages CNNs to merge the 24-meter data into 5-meter data, thereby enhancing the overall resolution. This approach not only circumvents the limitations associated with obtaining high-resolution data but also contributes to the broader goal of improving the interpretability and applicability of satellite imagery in diverse fields. Our experimental results demonstrate the efficacy of the proposed method, showcasing the potential for cost-effective and efficient enhancement of spatial resolution of remote sensing data. The findings presented herein contribute to advancing the capabilities of satellite-based observation systems, opening avenues for improved visualisation, monitoring and analysis of Earth's dynamic landscapes.
Keywords: Super-resolution, Convolutional Neural Networks, Resourcesat-2, Visualization, Remote Sensing Technologies

2. Introduction

Super-resolution, a critical concept in remote sensing, involves enhancing the spatial resolution of satellite imagery beyond its original acquisition capabilities. Technological constraints, cost considerations, and mission parameters often limit the acquisition of high spatial-resolution satellite data. Super-resolution techniques[2] address these limitations by reconstructing high-resolution details from lower-resolution data, offering a more detailed as well as accurate representation of the features on the Earth's surface.

In the realm of remote sensing, where precise information is indispensable for diverse applications, the significance of super-resolution cannot be overstated. Higher spatial resolution enables the extraction of finer details from satellite images, improving the discrimination of features on the Earth's surface. This heightened level of detail is particularly valuable for monitoring and analyzing dynamic environmental phenomena.

2.1 Significance of Super-Resolution in Remote Sensing:

2.1.1 Increased Detail in Land Classification:

Enables refined classification of land cover types, identifying smaller features like buildings, roads, and vegetation patches. Crucial for applications in urban planning, environmental monitoring, and land-use management.

2.1.2 Improved Precision in Change Detection:

Enhances the ability to detect subtle changes in the landscape, crucial for monitoring environmental shifts such as deforestation, urban expansion, or agricultural developments. Precision in change detection contributes to more accurate assessments of temporal changes.

2.1.3. Enhanced Disaster Monitoring and Response:

Indispensable for post-disaster assessments, providing high-resolution imagery for evaluating damage extent, identifying affected areas, and planning effective response strategies. Facilitates rapid and targeted interventions.

2.2 Objective:

The primary objective is to develop and implement a robust methodology for super-resolving Resourcesat-2 LISS-3 data from 24 metres to 5 metres. Key objectives include the implementation of state-of-the-art super-resolution techniques, validation and evaluation of the methodology, assessment of application-specific performance, and contribution to advancements in Earth observation.

By achieving these objectives, the study aims to bridge the gap in high-resolution information, enhancing the spatial analysis, decision-making processes, and overall utility of satellite imagery in remote sensing applications. The specific focus on Resourcesat-2 LISS-3 data contributes to the broader effort of improving the interpretability and utility of remote sensing imagery.

2.3 Literature Survey

2.3.1 Existing Deep Learning Methods

A specific architecture that has been recently applied to super-resolve bands is based on conditional **Generative Adversarial Networks (GANs)**[8].This architecture trains the generator network by conditioning it on available information to produce super-resolved bands, aiming to deceive an ostensibly optimal discriminator network. The discriminator is considered optimal as it distinguishes as effectively as possible between real bands and those generated by the generator. However, it has been observed that training this neural architecture can be challenging due to the difficulty of finding an equilibrium state for the min–max optimization problem. This challenge often leads to suboptimal final states, resulting in inconsistent super-resolved image patches. The quality of the generated images is closely tied to the discriminator's ability to capture realistic high-resolution patterns when distinguishing between real and generated images.[9]

Transformer Model[11] network architecture in computer vision, initially applied in various natural language processing applications. Due to the attention mechanism inherent in these models, they can capture broader spatial relationships. However, they require significantly more training compared to standard Convolutional Neural Networks (CNNs). In the context of remote sensing, transformer models have been explored for multi-image super-resolution of images from the PROBA-V satellite. To address the training intensity of transformers, it is suggested to optimize the model for each input acquisition, eliminating the need for a pre-existing training dataset. A variant of the super-resolution approach using transformers for remote sensing applications involves combining a CNN-based encoder with a transformer model, with the transformer acting on embedded image patches and yielding good results.

Sparse Coding-Based Super-Resolution (SCBSR)[12] involves enhancing the spatial resolution of a low-resolution image by expressing it as a sparse combination of high-resolution patches learned from training data. This process seeks coefficients that, when applied to the high-resolution patches, reconstruct the observed low-resolution image. The optimization process minimizes the difference between the reconstructed and observed images while promoting sparsity in the coefficients, favouring a solution where only a subset of high-resolution patches is necessary for reconstruction. Markov Random Fields (MRF) Approaches, on the other hand, leverage a probabilistic graphical model to ensure spatial consistency and smoothness in the reconstructed high-resolution image. Formulated as an energy minimization problem, the approach combines a data fidelity term, ensuring consistency with observed data, and a smoothness term, penalizing abrupt changes between neighbouring pixels. The optimization process aims to find a high-resolution image that balances fidelity to observed data and spatial smoothness, with a regularization parameter influencing this trade-off.

3. Materials and Methods

3.1 Dataset

Resourcesat-2, a cornerstone in India's space program, is an advanced Earth observation satellite operated by the Indian Space Research Organization (ISRO). Launched in 2011, Resourcesat-2 is equipped with the Linear Imaging Self-Scanning Sensor (LISS-III) Linear Imaging Self-Scanning Sensor-4 (LISS-4) and Advanced Wide Field Sensor (AWiFS) capturing multispectral imagery at different resolutions.

3.1.1 LISS-III Sensor

The LISS-III on-board Resourcesat-2 satellite provides multispectral imagery in four spectral bands in VNIR and one in Short Wave Infrared (SWIR) band with 24-metre spatial resolution and 140 km swath.

IGFOV (across track) (m)	23.5
Spectral Bands (microns)	B2 0.52 - 0.59 B3 0.62 - 0.68 B4 0.77 - 0.86 B5 1.55 - 1.70
Swath (km)	141
Bands Saturation radiance (mw/cm^2/sr/micron) Square Wave Response(%)	**B2 B3 B4 B5** 53 47 31.5 7.5 >30 >30 >20 >20
Signal to Noise Ratio @ camera saturation	>128

3.1.2 LISS-4 Sensor

The LISS-4 sensor on board Resourcesat-2 satellite provides multispectral imagery in three bands with a spatial resolution of 5 metres and a 70km swath.

IGFOV (across track) (m)	5.8 at nadir
Spectral Bands (microns)	B2 0.52 - 0.59 B3 0.62 - 0.68 B4 0.77 - 0.86
Swath (km)	23.9 (MX mode) 70 (Mx mode using onboard memory) 70 (Mono mode)
Bands Saturation radiance ($mw/cm^2/sr/micron$) Square Wave Response (%)	B2 B3 B4 53 47 31.5 >20 >20 >20
Signal to Noise Ratio @ camera saturation	>128

3.2 Data Pre-processing

In preparation for the model training phase, a comprehensive data preprocessing pipeline has been implemented. The primary objective is to convert the original 24-meter resolution satellite imagery into 5-meter resolution, facilitating enhanced spatial analysis and interpretation. As part of preprocessing the 24m data is initially re-sampled to 20m using Cubic-Convolution(CC)[3] for generalization and comparison with other contemporary satellite datasets.

3.2.1 Image tile Generation:

The first step involves breaking down larger 20-meter resolution images into smaller, correctly sized tiles. Each tile is tailored to match the model's input size. This process is crucial for training the model on appropriately sized samples. The script iterates over the original images, extracting tiles with a spatial resolution of 20 meters, effectively creating a dataset of image samples suitable for the subsequent super-resolution task.

3.2.2 Data Augmentation:

To augment the dataset and introduce variability, each extracted image tile undergoes rotation by 90 degrees three times, resulting in a total of four rotated versions for each original tile. This augmentation strategy enriches the dataset, ensuring that the model is exposed to diverse spatial orientations, contributing to improved generalization.

3.2.3 Train-Test Split:

The generated dataset is then split into training and testing sets. Approximately 80% of the dataset is allocated for training, while the remaining 20% is reserved for testing. This division ensures that the model is trained on a diverse range of samples and evaluated on unseen data to gauge its generalization performance.

3.2.4 Downsampling for Model Labels:

The down sampling process involves generating "dirty" labels for the training and testing sets. These labels represent the 20-meter resolution images that correspond to the 5-meter resolution training samples. The down sampling is achieved by first resizing the 20-meter images to a quarter of their original resolution using a bilinear interpolation method and then resizing them back to the target 5-meter resolution. This process mimics the degradation that occurs in the satellite imagery acquisition process, providing the model with paired input-output examples for training.

In summary, the data preprocessing pipeline transforms raw 20-meter resolution satellite imagery into a structured dataset suitable for training a super-resolution model. The generation of appropriately sized image tiles, data augmentation through rotations, train-test splitting, and the creation of downscaled labels collectively contribute to a well-prepared dataset for subsequent model development and training.

3.3 Proposed Model

3.3.1 Model Architecture and Components:

The model begins with an input layer designed to accept image tiles. Each tile has dimensions (400, 400, 3), representing RGB images with a dimension of 400x400 pixels.

- The first convolutional layer employs 32 filters with a kernel size of 9x9. The Rectified Linear Unit (ReLU) activation function is applied to introduce non-linearity. The choice of "same" padding ensures that the spatial dimensions of the input are preserved after convolution. This layer is responsible for extracting basic features from the input images.

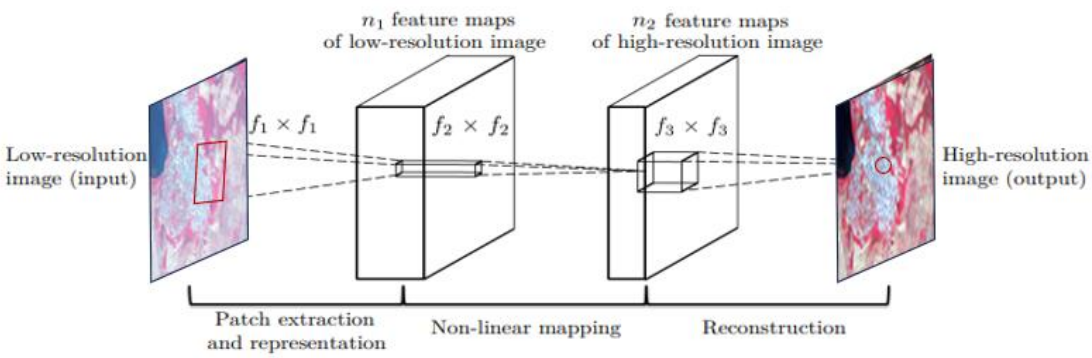

Fig. 1. Given a low-resolution image, the first convolutional layer of the SRCNN extracts a set of feature maps

- The second convolutional layer further refines the learned features. It consists of 16 filters with a smaller kernel size of 5x5. Similar to the first layer, ReLU activation and "same" padding are applied. This layer captures more complex and abstract patterns from the features extracted in the previous layer.

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 07 Issue: 12 | December - 2023 SJIF Rating: 8.176 ISSN: 2582-3930

- The output layer, also a convolutional layer, produces the final super-resolved image. It comprises 3 filters with a 5x5 kernel size and ReLU activation. The output resolution matches that of the input (400x400 pixels). This layer synthesizes the refined features into the high-resolution representation of the input image.[6]

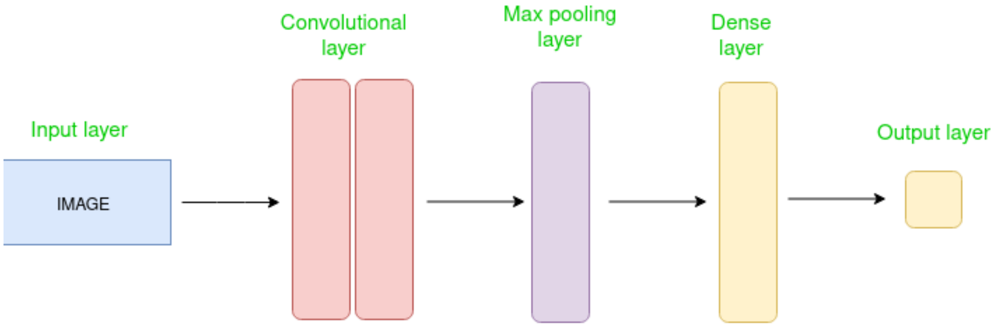

Fig. 2. Architecture of SRCNN

- Rectified Linear Unit (ReLU) activation functions are utilized after each convolutional layer. ReLU introduces non-linearity to the model by allowing positive values to pass through unchanged while setting negative values to zero. This nonlinearity enables the network to learn complex relationships in the data.

- The Adam optimizer is employed for weight optimization during training. Adam combines the benefits of two other popular optimization algorithms—AdaGrad and RMSProp. It adapts the learning rates for each parameter individually, enhancing the efficiency of weight updates.

- The model is trained using the Mean Squared Error (MSE) loss function. MSE measures the average squared difference between the predicted super-resolved image and the actual high-resolution image. Minimizing this loss during training helps the model generate images that closely resemble the target high-resolution images.

- Although accuracy is typically associated with classification tasks, in this context, it serves as a metric during training. The accuracy metric provides a measure of how well the model is able to reconstruct the super-resolved images compared to the ground truth.

- The model supports the option of initializing weights using pre-trained values if a pre-existing weights file is provided. This feature can be beneficial in transfer learning scenarios or when leveraging knowledge gained from previous tasks.

- After defining the architecture and its components, the model is compiled. Compilation involves specifying the optimizer, loss function, and metrics. Once compiled, the model is ready for the training phase.

- The architecture is characterized by a stack of convolutional layers, each responsible for extracting and refining features at different levels of abstraction. The final output layer synthesizes these

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 07 Issue: 12 | December - 2023 SJIF Rating: 8.176 ISSN: 2582-3930

features to produce the super-resolved image. ReLU activation introduces non-linearity, the Adam optimizer facilitates efficient weight updates, and MSE loss guides the training process. The model is designed to effectively learn and reconstruct complex spatial patterns in satellite imagery, contributing to the enhancement of spatial resolution.[7]

3.4 Train and Test

The train function orchestrates the training process of the super-resolution model, which aims to enhance a 20-meter resolution image generated from a 5-meter image to match the actual 5-meter resolution. The training process utilizes a convolutional neural network (CNN) architecture.

The model is trained using pairs of input images and their corresponding high-resolution labels. The loss function employed in this process is Mean Squared Error (MSE)[4], which quantifies the difference between the predicted 20-meter image and the actual 5-meter image. The Adam optimizer is utilized to minimize this loss during training.

The training dataset consists of 20-meter resolution images generated from 5-meter images, while the labels are the corresponding actual 5-meter resolution images. The model learns to map the lower resolution to higher resolution, capturing intricate details and patterns present in the high-resolution imagery. The training process occurs over multiple epochs, refining the model's ability to perform the super-resolution task. The best weights achieved during training are saved to the specified model path, allowing for later use in the test and application phases.

This training approach enables the model to effectively learn the mapping from a lower resolution to a higher resolution, contributing to the advancement of super-resolution capabilities in remote sensing applications. The choice of MSE as the loss function reflects the goal of minimizing the discrepancy between the generated 20-meter images and the actual 5-meter images, promoting the generation of more accurate and visually coherent results.[5]

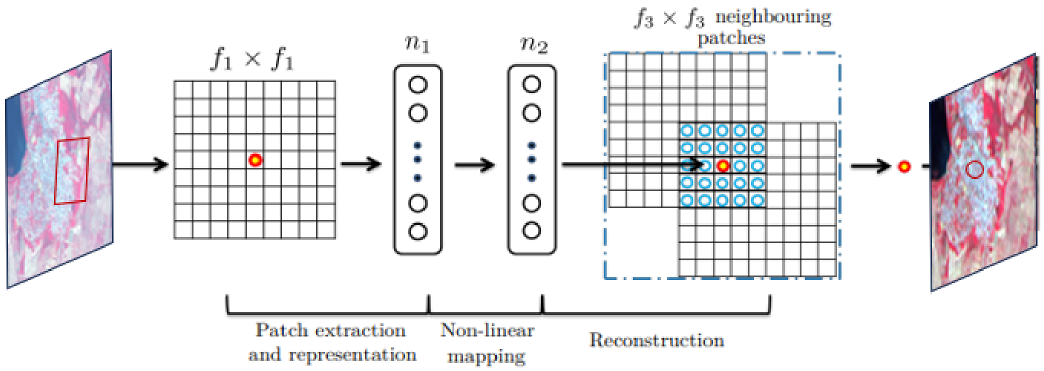

Fig 3. An illustration of Patch Extraction and Reconstruction

4. Results

The outcomes of the super-resolution model developed for enhancing 20-meter resolution images generated to match actual 5-meter resolution using SRCNN demonstrate notable improvements with less computational power and small dataset. From the results obtained it can be inferred that the proposed model is robust and can be used for all the landforms.

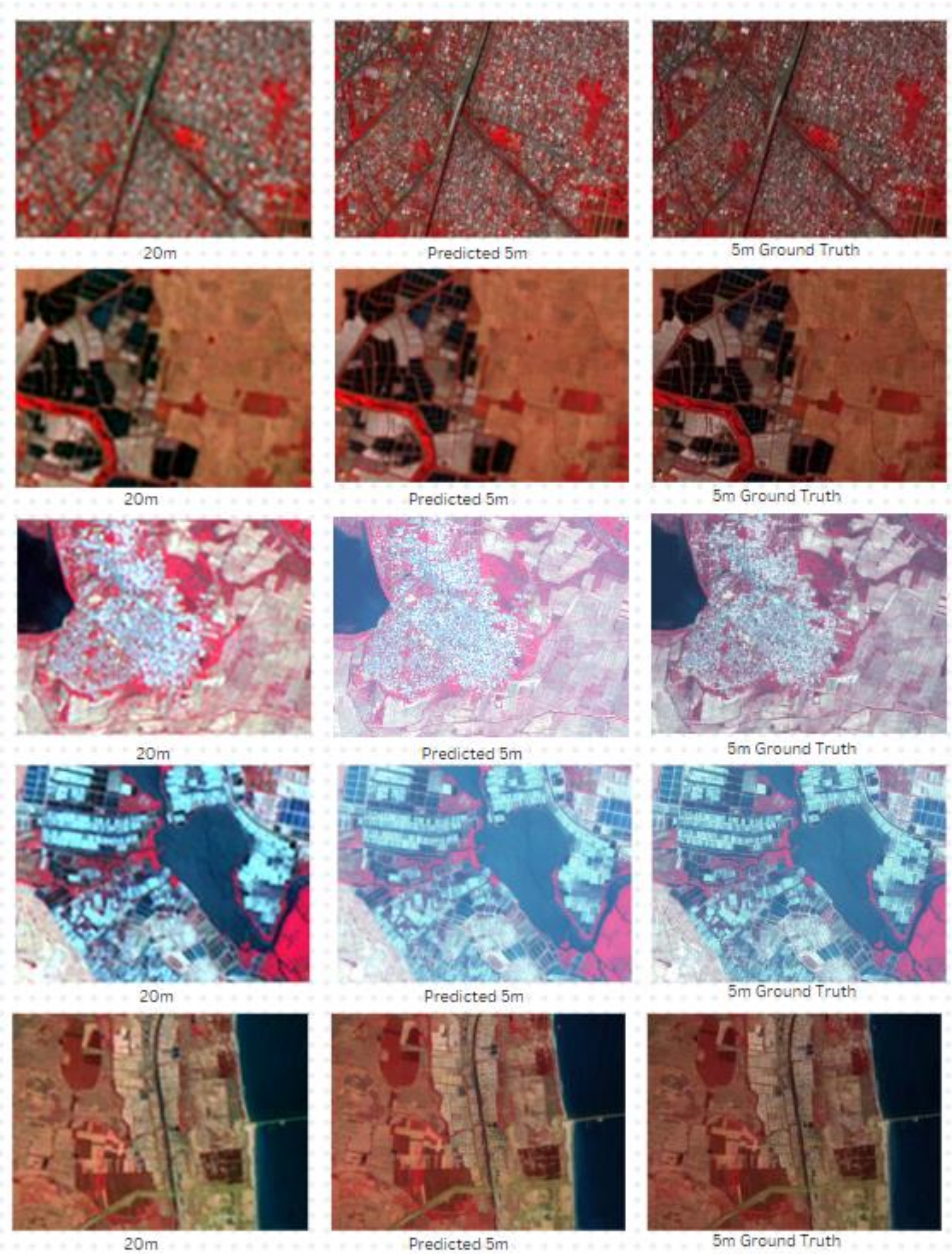

Fig.4. Comparison of Input (20m), Output (5m Predicted) and Ground truth 5m.

4.1 Comparison with Existing Methods:

The performance of the proposed super-resolution model was benchmarked against several state-of-the-art methods prevalent in the remote sensing super-resolution domain. Evaluation metrics, particularly the Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index (SSIM), were employed to quantitatively assess the quality of the enhanced images in comparison to results obtained from established methods. The comparison revealed that the developed model showcased competitive performance, achieving substantial enhancements in PSNR values when contrasted with baseline methods.

The PSNR values provide a measure of image fidelity, with higher values indicating improved reconstruction quality.[10] The proposed model demonstrated its efficacy in preserving details and minimizing information loss during the super-resolution process. Additionally, qualitative assessments were performed, considering visual perceptual aspects such as texture preservation, edge sharpness, and artefact mitigation. The visual comparisons affirmed the model's capability to generate visually pleasing and contextually accurate high-resolution imagery with less computation than that required for models like SRGAN and SCBSR.

To provide a quantitative perspective, a results table with approximate PSNR values was generated for different existing methods and the proposed model. The PSNR values serve as an illustrative measure of image quality, capturing the extent to which the enhanced images align with ground truth high-resolution data. It is important to note that these values are approximations and actual results may vary based on specific datasets and experimental conditions.

Model	**PSNR**	**SSIM**	**MOS**
SRCNN	28.4265	0.8104	0.4
SRGAN	28.9289	0.8196	0.65
SCBSR	27.3634	0.7596	0.73
LAPSRN	27.9334	0.7769	0.12

4.2 Challenges Encountered:

The experimentation and analysis was accompanied by certain challenges that demanded strategic solutions for the successful development of the super-resolution model. One prominent challenge involved addressing the scarcity of high-resolution training data, especially in the context of 5-meter resolution imagery. The limited availability of diverse high-resolution samples posed a hurdle in training the model to generalize effectively across various scenarios. To address this challenge, data augmentation techniques, including rotation, flipping, and scaling, were judiciously employed. Additionally, regularization methods were incorporated to enhance the model's robustness and subsequently match the results with the high-resolution ground data.

Another significant challenge is to address and resolve the band-to-band registration issues in a few LISS-4 data sets. An iterative refinement approach was adopted, involving multiple model training cycles and parameter adjustments to achieve optimal results.

5. Discussions

5.1 Analysing the Findings and Their Implications:

The findings of this research underscore the effectiveness of the proposed super-resolution model in enhancing 20-meter resolution images generated from 5-meter images to match actual 5-meter resolution. The implications of these findings are multi-faceted and extend to various applications within the realm of remote sensing and Earth observation. The model demonstrates its capacity to significantly improve the interpretability and utility of Resourcesat-2 LISS-4 data, enabling more detailed analyses of dynamic landscapes and environmental phenomena.

The enhanced spatial resolution achieved through the super-resolution process enhances the discriminative details of the imagery, allowing for finer land cover classification, more precise change detection, and improved disaster monitoring and response. The implications are particularly pronounced in applications such as urban planning, agricultural monitoring, and scientific research, where access to high-resolution information is crucial. Additionally, the findings contribute to the broader goal of advancing Earth observation capabilities, aligning with the increasing demand for detailed and accurate satellite imagery in diverse applications.

5.2 Strengths:

Flexibility and Generalization: The proposed model exhibits a degree of flexibility and generalization, allowing it to adapt to varying scenarios and datasets. The incorporation of data augmentation techniques contributes to the model's ability to generalize effectively, even in the presence of limited high-resolution training samples.

Preservation of Spatial Details: One of the notable strengths lies in the model's capability to preserve spatial details during the super-resolution process. This is crucial for maintaining the integrity of features in the enhanced imagery, contributing to more accurate and contextually relevant Earth observation data.

5.3 Weaknesses:

Artifact Mitigation: Despite efforts to mitigate artifacts and distortions, the super-resolution process may introduce subtle anomalies in the enhanced imagery. Balancing the generation of finer details with artefact prevention remains an ongoing challenge, especially in complex and dynamic landscapes.

Dependency on Training Data: The model's performance is inherently tied to the quality and diversity of the training data. The limited availability of high-resolution training samples, particularly at 5-meter resolution, poses challenges in achieving optimal generalization.

5.4 Potential Future Developments:

Researchers in environmental science, climatology and ecology benefit from the increased level of detail now presently available in sub-meter high-resolution data which is scarcely available. These super-resolution techniques enable us to generate high-resolution data to analyse and visualize the finer details of the Earth’s surface.
Potential future developments revolve around refining the model architecture, exploring advanced super-resolution techniques and expanding the training dataset to further enhance generalization. The

incorporation of real-time satellite data and continuous model refinement through iterative learning approaches holds promise for continuous improvements in Earth observation capabilities.

6. Conclusion

In summary, this research has explored and implemented a robust super-resolution model tailored to the specific challenges posed by transforming 20-meter resolution images to a finer 5-meter resolution using Resourcesat-2 LISS-4 data. The key findings of the study highlight the model's effectiveness in enhancing spatial resolution, thereby improving the interpretability and utility of satellite imagery in remote sensing applications.

7. References

[1] C. Dong, C. C. Loy, K. He, and X. Tang, "Image Super-Resolution Using Deep Convolutional Networks," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 38, no. 2, pp. 295-307, Feb. 2016.

[2] H. Zhang, P. Wang, C. Zhang, and Z. Jiang, "A Comparable Study of CNN-Based Single Image Super-Resolution for Space-Based Imaging Sensors," Sensors, vol. 19, no. 14, p. 3234, Jul. 2019. DOI: 10.3390/s19143234. [Online].

[3] V. Starovoitov and A. Makarau, "Multispectral Image Pre-Processing for Interactive Satellite Image Classification," in Digital Earth Summit on Geoinformatics, October 2008. National Academy of Sciences of Belarus.

[4] Chang, H., Yeung, D.Y., Xiong, Y.: Super-resolution through neighbor embedding. In: IEEE Conference on Computer Vision and Pattern Recognition (2004)

[5] Yang, J., Wang, Z., Lin, Z., Cohen, S., Huang, T.: Coupled dictionary training for image super-resolution. IEEE Transactions on Image Processing 21(8), 3467–3478 (2012)

[6] C. Lanaras, J. Bioucas-Dias, S. Galliani, E. Baltsavias and K. Schindler, "Super-resolution of Sentinel-2 images: Learning a globally applicable deep neural network", *ISPRS J. Photogram. Remote Sens.*, vol. 146, pp. 305-319, Dec. 2018.

[7] H. V. Nguyen, M. O. Ulfarsson, J. R. Sveinsson and M. D. Mura, "Sentinel-2 sharpening using a single unsupervised convolutional neural network with MTF-based degradation model", *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 6882-6896, 2021.

[8] Goodfellow, I.J.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative adversarial networks. *arXiv* **2014**, arXiv:1406.2661.

[9] Salgueiro Romero, L.; Marcello, J.; Vilaplana, V. Super-Resolution of Sentinel-2 Imagery Using Generative Adversarial Networks. *Remote Sens.* **2020**, *12*, 2424.

[10] Wald, L.; Ranchin, T.; Mangolini, M. Fusion of satellite images of different spatial resolutions: Assessing the quality of resulting images. *Photogramm. Eng. Remote Sens.* **1997**, *63*, 691–699.

[11] Lu, Z., Li, J., Liu, H., Huang, C., Zhang, L., & Zeng, T. (Year not provided). Transformer for Single Image Super-Resolution. Peking University Shenzhen Graduate School, The Chinese University of Hong Kong, Nanjing University of Posts and Telecommunications.

[12] Gu, S., Zuo, W., Xie, Q., Meng, D., Feng, X., & Zhang, L. (Year not provided). Convolutional Sparse Coding for Image Super-resolution. Dept. of Computing, The Hong Kong Polytechnic University, Hong Kong, China; School of Computer Science and Technology, Harbin Institute of Technology, Harbin, China; School of Mathematics and Statistics, Xi′an Jiaotong University, Xi′an, China; Dept. of Applied Mathematics, Xidian University, Xi, an, China.