DOI: 10.6977/IJoSI.202508_9(4).0004
J. Jayanth, H. K. Ravikiran, etc./Int. J. Systematic Innovation, 9(4), 42-53 (2025)

Single-frame super-resolution with deep residual network–generative adversarial networks

J. Jayanth[1], H. K. Ravikiran[2]*, T. Yuvaraju[1], R. Dileep[1]

[1]Department of Electronics and Communication Engineering, GSSS Institute of Engineering and Technology for Women, Visvesvaraya Technological University, Belagavi, Karnataka, India

[2]Department of Electronics and Communication Engineering, Navkis College of Engineering, Visvesvaraya Technological University, Belagavi, Karnataka, India

*Corresponding author E-mail: hkr@navkisce.ac.in

(Received 14 February, 2025; Final version received 24 April, 2025; Accepted 16 May, 2025)

Abstract

Developing and evaluating a deep learning-based method to enhance satellite image resolution has emerged as a promising approach to address challenges posed by motion, imaging blur, and noise without modifying existing optical systems. This study utilized an enhanced super-resolution generative adversarial network (SRGAN) with ResNet-50 as the generator and a modified VGG-19 in the discriminator. The model was trained on remote sensing images from the Linear Imaging Self-Scanning imagery and compared with very deep super resolution, SRGAN, and enhanced SRGAN methods using the structural similarity index measure (SSIM) and peak signal-to-noise ratio (PSNR) as evaluation metrics. Utilizing an enhanced SRGAN with ResNet-50 and modified VGG-19 significantly improved satellite image resolution. The proposed method consistently outperformed conventional convolutional neural network- and generative adversarial network-based super-resolution techniques. Across three test datasets, the method achieved SSIM scores as high as 0.862 and PSNR scores of 33.256, 32.886, and 34.885, demonstrating its superior ability to preserve image properties and enhance resolution. The incorporation of perceptual loss alongside pixel loss contributed to improved visual quality, making the approach particularly effective in maintaining fine details and naturalistic high-frequency characteristics.

Keywords: Generative Adversarial Network, Linear Imaging Self-Scanning Image, Peak Signal-to-Noise Ratio, ResNet-50, Structural Similarity Index Measure, Super Resolution

1. Introduction

Super-resolution (SR) reconstruction is a technology that generates high-resolution (HR) images from a sequence of low-resolution (LR) images using a particular algorithm, all without altering the settings of the imaging hardware (Adarsh et al., 2020). Ultra-high-definition television, emergency event monitoring, target identification and localization, military precision targeting, battlefield environment surveillance, and medical image diagnosis are just a few of the many applications of image SR reconstruction technology in both civilian and military sectors (Wang et al., 2022).

Spatial resolution, the lowest discernible unit size or dimension in remote sensing images, is a measure of the image's ability to identify ground target features (Han et al., 2023). Finer target identification is made possible by higher spatial resolution, suggesting that remote sensing photographs include a greater amount of information about ground objects. However, attaining higher spatial resolution solely through hardware advancements is difficult due to the downsampling effects of imaging sensors and various factors that cause deterioration in satellite image processing. These difficulties result in high development costs and lengthy hardware iteration cycles.

Nonetheless, by reconstructing HR images from LR but easily accessible images, image SR technology offers an affordable means to acquire HR images (Sui et al., 2023). Conventional SR reconstruction approaches, including interpolation

and prior information-based reconstruction, possess limitations. Interpolation techniques, such as bilinear and bicubic interpolation, perform well in real time but suffer from observable edge effects and poor detail recovery performance (Sui et al., 2023; Wang et al., 2023; Zhang et al., 2022). Reconstruction techniques based on prior knowledge often rely on constraints such as maximal a posteriori probability and iterative back-projection; however, these methods are computationally costly, have limited applicability, and exhibit poor generalizability (Bu et al., 2024; Dong et al., 2022; Haut et al., 2018).

As deep learning techniques advance, single-image SR techniques based on them outperform conventional methods in remote sensing applications and demonstrate wide-ranging utility (Khan et al., 2023; Pang et al., 2023). For example, the SR convolutional neural network (SRCNN) model applies non-linear mapping to extract LR image features for reconstruction (Wang et al., 2023). However, SRCNN's limitations include its single-scale applicability, slow training convergence, and shortcomings in handling various image contents. To overcome these challenges, W. This innovative approach significantly enhances the model's adaptability, training speed, and capacity to capture intricate image nuances, addressing the shortcomings of SRCNN effectively.

The advent of VGG networks has led to the widespread use of deeper network models (Shi et al., 2023). Residual networks (ResNet) (Wang et al., 2023), offer a solution to the gradient-vanishing issues in deep networks. A very deep SR (VDSR) model is used by incorporating ResNet to extract high-frequency information residuals and capture more image detail information (Frizza et al., 2022). Further innovations, such as multiscale ResNet (Zhang et al., 2022), cascading ResNet, and enhanced ResNet (Liu et al., 2022), fully utilize low-level features with multiscale residual blocks for feature extraction and fusion.

The residual channel attention network (Tang et al., 2022) is introduced as a mean of integrating attention processes into the SR field, and the generative adversarial networks (GANs) (Wang et al., 2021) are developed through advancements in deep learning. To address issues with excessively smooth reconstructed images, a SR generative adversarial network (SRGAN), which introduced perceptual loss and utilized the GAN framework for adversarial training, was introduced (Zhang et al., 2021). The more realistic textures produced by enhanced SRGAN (ESRGAN) outperformed those of SRGAN (Wang et al., 2024). Several algorithms, including SRGAN (Min et al., 2024), extended regularized adversarial GAN (Lei et al. 2017), and transferred GAN (He et al., 2022; Jiang et al., 2018; Li et al., 2022; Wang et al., 2022; Veganzones et al., 2016), have been developed for use in remote sensing applications. These algorithms combine multi-loss training techniques, semantic segmentation, and attention mechanisms to enhance the reconstruction of remote sensing images. Furthermore, research on resolution enhancement extends to other types of remote sensing images, including multisource image fusion (Dileep et al., 2024; Jayanth et al., 2025) and multispectral imaging (Ravikiran et al., 2024).

Recent advancements in unsupervised hyperspectral image SR have significantly enhanced the reconstruction of HR images without requiring paired training data. Li et al. (2025) proposed an enhanced deep image prior network that leverages network structure as an implicit prior to guide reconstruction. The model-informed multistage unsupervised network integrates degradation models into a multistage architecture for progressive refinement (Li et al., 2024). Meanwhile, the X-shaped interactive autoencoders utilize cross-modality mutual learning between spectral and spatial domains, while the model-guided coarse-to-fine fusion network applies a coarse-to-fine reconstruction strategy guided by physical priors. These approaches contribute to improved spatial-spectral fidelity in real-world applications.

Despite the remarkable achievements of GANs in image reconstruction and style transfer, challenges persist in their training process, such as mode collapse and gradient vanishing. Moreover, the majority of current techniques use pixel-level loss functions, such as mean squared error (MSE), potentially resulting in excessively smooth reconstructed images devoid of high-frequency information. Given the intricacy of scenes and the variety of target attributes in remote sensing photographs, the qualities of actual remote sensing datasets must be considered throughout the reconstruction process.

The generator in our single-frame SRGAN framework is designed using the ResNet-50 architecture, chosen for its strength in training deep networks via residual learning. This helps mitigate issues such as vanishing gradients and supports the learning of complex, high-level features. In this implementation, the generator receives an LR satellite image and a random noise vector as inputs. These are concatenated to form a rich input space that combines deterministic and stochastic components. This fusion enables the network to produce more realistic and varied HR images (Dileep et al., 2025). The core of the generator consists of several residual blocks, each containing convolutional layers, batch normalization, and rectified linear unit (ReLU) activations. The skip connections in these blocks facilitate the retention of low-level features while capturing high-level abstractions, helping to preserve spatial and textural details (Li et al., 2023).

Following feature extraction, the generator uses upsampling layers—typically transposed convolutions—to increase the spatial resolution of the feature maps. The final output is produced through convolutional layers that refine these upsampled features into a coherent HR image. On the other hand, a discriminator is built upon a modified VGG-19 architecture, which is well-regarded for its effectiveness in perceptual feature extraction. The VGG-19 network is pre-trained and adapted to serve as a binary classifier within the GAN framework, distinguishing between real HR images and those generated by the network. Its deep yet straightforward structure makes it ideal for identifying fine-grained visual differences, thereby enhancing the adversarial training process and helping the generator improve the realism of its outputs (Li et al., 2023).

In this work, the input image passes through a series of convolutional layers with ReLU activations, interleaved with max pooling operations to progressively extract higher-level features. The deep convolutional structure helps the discriminator capture fine-grained texture patterns and global structures, both of which are essential for assessing image realism. After convolutional feature extraction, the output is flattened and passed through multiple dense layers with *LeakyReLU* activations, allowing the network to model more complex decision boundaries. The final layer uses a sigmoid function to produce a probability score indicating whether the input image is real or generated.

By combining the perceptual strength of VGG-19 with a GAN setup, the discriminator plays a crucial role in guiding the generator to produce visually convincing HR images. This adversarial training setup encourages the generator to create outputs that are not only structurally accurate but also perceptually indistinguishable from real images.

In conventional image SR tasks, loss functions, such as MSE and mean absolute error, are commonly used to measure the pixel-wise differences between the generated and ground truth images. However, while these losses are mathematically straightforward and encourage the generated image to be numerically close to the target, they often lead to overly smooth outputs that lack high-frequency details and perceptual sharpness, especially in complex images, such as satellite imagery, where texture and structure are critical.

To address this limitation, perceptual loss is introduced. Unlike conventional pixel-level loss functions, perceptual loss evaluates the difference between high-level feature representations of the generated and ground truth images, extracted from a pre-trained deep convolutional network (typically a network like VGG-19). These feature maps capture semantic content, texture, and structural patterns that align more closely with how humans perceive image quality.

By minimizing perceptual loss, the generator is guided to produce images that are not only numerically similar but also visually and structurally closer to real HR images. This results in outputs that retain finer details, sharper edges, and more natural textures.

Furthermore, current SR techniques may not generalize well to new targets and situations, necessitating training methods and model designs that explicitly improve resilience and generalization. To address these challenges, modifications to the GAN were implemented, specifically by adding ResNet-50 in the generator network and VGG-19 in the discriminator model. The modifications are as follows:

- To enhance channel efficiency, the generator network was modified by switching out its basic residual blocks with bottleneck blocks from ResNet-50 and extracting the features that are important for SR
- Improved skip connections make information flow between layers more effective and aid in the reconstruction of HR features. In order to improve feature blending and guarantee reliable classifier performance, an additional fully connected (FC) layer was implemented
- Effective discriminative feature extraction was improved by fine-tuning VGG-19 within the discriminator.

The article's structure is as follows: Section 2 outlines the methodology, detailing the GAN; Section 3 presents the results and discussion along with a comparison with other algorithms using different datasets; finally, Section 4 provides a brief conclusion.

2. Methodology

In SR, GANs have become a dominant approach, providing a strong framework for converting LR images into HR images. The SRGAN architecture typically consists of a generator and a discriminator, both are essential components in the creation and assessment of super-resolved images.

2.1. General SRGAN Architecture

The algorithm for the SRGAN is as follows:

(i) Input: LR image
(ii) Architecture: Series of convolutional layers without multiscale residual blocks, followed by batch normalization and activation functions
(iii) Upsampling: The generator aims to learn the mapping from an LR image to an SR image

(iv) Output: SR image with dimensions rW × rH × C, where *r* is the upsampling factor
(v) The complete generator function is:

G(LR image) = SR image.

When the generator network transitioned to ResNet-50, the bottleneck blocks in this method generator model took the place of the basic residual blocks. The discriminator model evolved concurrently into a condensed version of the VGG-19 network, retaining the first 16 layers and including an FC layer for enhanced functionality. Bottleneck blocks were used in the architecture, which used ResNet-50, to improve performance. Lowering the number of channels and increasing processing performance in deeper networks were made possible in large part by the bottleneck blocks. This design decision was crucial for handling the computing power requirements of deep network architectures.

The generator based on ResNet-50 comprised an input layer, ResNet blocks for feature extraction, upsampling layers to improve spatial resolution, and a final output layer that generates the HR image. Batch normalization, ReLU activation, and convolutional layers were combined to form the ResNet blocks. This study used bottleneck blocks to improve speed while generating SR using ResNet-50. Remarkably, when the stride was set to 1, neither the basic nor the bottleneck blocks performed downsampling. However, when the stride was set to 2, downsampling occurred prior to the addition operation. The main purpose of the convolution kernel was to reduce the number of channels that are handled later in the module. With fewer parameters in the channel, this reduction maximized its use.

A pooling layer was layered following a conventional convolutional layer in the first construction layer. The three remaining bottleneck blocks were then integrated into the second construction tier. Downsampling residual bottleneck blocks was the first step in the third, fourth, and fifth construction layers, followed respectively by three, five, and two residual bottleneck blocks. After processing 16 bottleneck blocks for feature extraction, pixel-shuffle layers were used to improve the resolution twice. Crucially, this optimization happened deliberately at the end of the network layers, helping to maximize resolution while reducing the amount of processing resources used.

The architecture of the generator was carefully designed to learn the complex mapping from LR images to SR images. The generator was made up of a sequence of convolutional layers without multiscale residual blocks. It also included activation functions and batch normalization. Increasing the spatial resolution of LR images was the main goal, and an SR image with dimensions rW × rH × C—where *r* denotes the upsampling factor—is the result. The discriminator functioned to separate the generated SR images from genuine HR images at the same time. Convolutional layers were used in conjunction with batch normalization and activation functions to generate a binary classification that indicates the likelihood that an image is a true HR representation.

A wide range of loss functions is necessary for the SRGAN to perform well. By measuring the generator's realism, the adversarial loss (*advLadv*) creates a competitive environment between the generator and discriminator. Perceptual dissimilarity between produced and HR images was measured by the perceptual loss (*Lperceptual*), ensuring that important features are retained. Furthermore, the pixel-by-pixel variations were quantified by the MSE, which offers a more detailed evaluation of fidelity.

A complex feature extraction technique was integrated into the generator's design using ResNet-50. The LR satellite image was represented by *Ilow* in the input layer of the latent space, while *z* provided a random noise vector. The first input, *x0*, was formed by concatenating *Ilow* and *z*. To extract features, a set of ResNet blocks—each consisting of convolutional layers, batch normalization, and ReLU activation—was combined. The ResNet blocks' skip connections promoted gradient flow, reducing the effects of disappearing gradients. The next step involved an upsampling layer that used transposed convolutions or other methods to improve spatial resolution. The final output layer, which was composed of convolutional layers with a suitable activation function (*tanh*), yielded the HR image *I^high*. This output resulted from the upsampled features, represented as *xupsampled*.

The residual learning paradigm was embodied by the ResNet block, which was defined as $x + Convolution(ReLU(BatchNorm(Convolution(x))))$. The trainable parameters *þG*, representing the weights of the convolutional layers, were encapsulated in the generator function as a whole, represented as *G(Ilow; θG) = I^high*. Essentially, the proposed SRGAN architecture incorporates ResNet-50's ability to smoothly capture complex features, offering a strong foundation for producing realistic and detailed HR satellite images. The design advances the state-of-the-art in SR image synthesis by achieving a delicate balance through the complex interplay of the generator, discriminator, and the stated loss functions.

2.2. Generator Architecture with ResNet-50

The algorithm for the generator architecture is as follows:

(i) Input layer (latent space): *Ilow* is the input LR satellite image, and *z* is a random noise vector.
- $x_0 = Concatenate(Ilow, z)$

(ii) ResNet blocks:
- Integrate several ResNet blocks for feature extraction.
- Each ResNet block consists of convolutional layers, batch normalization, and ReLU activation.
- The skip connections help in gradient flow.
- $X_i = ResNetBlock(x_i - 1)$, $i = 1, 2, \ldots, n$.

(iii) Upsampling layers:
- Upsample the features to increase spatial resolution.
- Use transposed convolutions or other upsampling techniques.
- *Xupsampled = Upsample(xn)*

(iv) Output layer: Produce the final HR image using convolutional layers with a suitable activation function (e.g., *tanh*).
- *I^high = Convolution(xupsampled)*

(v) Each ResNet block can be defined as follows: *ResNetBlock(x) = x + Convolution(ReLU(Batch Norm(Convolution(x))))*

where *Convolution* refers to a 3 × 3 convolutional layer, *BatchNorm* is batch normalization, and *ReLU* is the ReLU activation function.

(vi) Complete generator function:
$G(Ilow; \theta_G) = I^{high}$

Where θ_G represents the weights of the convolutional layers in the generator.

This architecture leverages ResNet-50's ability to capture intricate features, helping the generator produce more realistic and detailed HR satellite images.

2.3. Discriminator with VGG-19

The algorithm for the discriminator with VGG-19 is as follows:

(i) Input: Real HR images (I_{HR}) and generated SR images (I_{SR}).

(ii) Architecture: Convolutional layers followed by batch normalization and activation functions.

(iii) Objective: Discriminate between real and generated images.

(iv) Output: Binary classification indicating the probability of being a real HR image.

(v) Complete discriminator function:

(high)→Binary ClassificationD(Ihigh; θD)→ Binary Classification

With a VGG-19 architecture, the discriminator functions as a discriminating agent in the GAN, especially for SR tasks. This discriminator is equipped with a complex architecture inspired by VGG-19 and incorporates an advanced method to discriminate between generated SR and genuine HR images.

The discriminator performs feature extraction, discriminative analysis, and classification, leveraging the VGG-19 architecture as its base. The discriminator consists of six block structures, each composed of convolutional and FC layers, enabling feature extraction and classification across 19 layers. Most importantly, it uses the first 16 layers to generate feature maps that capture important details from the input images.

After an image is input into the discriminator, it undergoes a transformation process involving discriminative analysis and feature extraction. Initially, the image passes through the first convolution block, which utilizes a 3 × 3 kernel size and a stride. To reduce computational complexity and prevent overfitting, the convolved image is downsampled using max pooling procedures.

This process progresses through subsequent blocks corresponding to the second, third, fourth, and fifth phases of the network, where deeper convolutions hone and enhance the characteristics that have been recovered. When the image reaches the FC layers, it experiences a significant metamorphosis into a vector. The next step involves passing this vector across thick layers, each with 1,024 neural units. Here, the discriminator's capacity to distinguish minute variations between generated and actual images is improved by the addition of non-linearity to its decision-making process using the *LeakyReLU* activation function. As the image propagates through the discriminator, it is subjected to a number of adjustments to extract complex characteristics and identify minute distinctions between generated SR and genuine HR images. The discriminator examines textures, structures, and general appearances using max pooling and convolutional procedures to cultivate a robust discriminative capability.

When classification reaches its final stage, the discriminator thoroughly assesses the converted vector. The last dense layer has an output dimension of two, denoting true or false depending on the classification outcomes. A sigmoid activation function is applied to produce a probability score for each image, classifying the image as either HR or SR. Notably, the sigmoid layer is preceded by an extra-thick FC7 layer that improves the discrimination between SR images and ground truth images, as well as enhances the discriminator's capacity to aggregate features from HR images.

2.4. Discriminator with VGG-19

The algorithm for the discriminator with VGG-19 is as follows:

(i) Input: LR or SR image (*I*).

(ii) Initialization:

- Load pre-trained modified VGG-19 model weights.
- Extract the convolutional and FC layers for feature extraction and classification.

(iii) Feature extraction:
- Pass the input image through the first convolution block:
  (a) *F1 = Conv(I, W1)*
  (b) *A1 = ReLU(F1)*
- Perform max pooling:
  P1 = MaxPool(A1)
- Iterate through subsequent convolution blocks.
- Apply convolutions and activations:
  (a) *Fi = Conv(Pi − 1, Wi)*
  (b) *Ai = ReLU(Fi)*
- Utilize max pooling:
  Pi = MaxPool(Ai)

(iv) Dense layer transformation:
- Flatten the feature maps into a vector:
  V = Flatten(Pn)
- Pass the vector through dense layers:
  (a) *Z1 = V·Wfc1 + bfc1*
  (b) *Afc1 = LeakyReLU(Z1)*
  (c) *Z2 = Afc1·Wfc2 + bfc2*
  (d) *Afc2 = LeakyReLU(Z2)*

(v) Classification: Pass the transformed vector through the final dense layer:
  (a) *Z3 = Afc2·Wfc3 + bfc3*
  (b) *Afc3 = Sigmoid(Z3)*

(vi) Output: Probability score indicating the authenticity of the input image as either LR or SR.

2.5. Loss Function

The two primary components of the loss for single-image SR (*lSSR*) are the regularization loss and the perceptual loss.

(i) Perceptual loss (*lperceptual*): In a perceptual space, perceptual loss quantifies the difference between the generated SR images and the HR ground truth images. It is frequently calculated using a feature similarity metric between the feature representations of the HR and SR images taken from a pre-trained deep neural network, such as the MSE or cosine similarity. The perceptual loss may be mathematically represented as follows:

$$lperceptual = \frac{1}{N}\sum_{i=1}^{N}\left\|\Phi(I_{HR}^{i}) - \Phi\left(G\left(I_{LR}^{i}\right)\right)\right\|_2^2 \quad (1)$$

Where N is the total number of training samples, G stands for the SR generator model, ϕ for the feature extractor, and $I_{HR}^{(i)}$ and $I_{LR}^{(i)}$ for the *i-th* HR and LR input images, respectively.

(ii) Regularization loss (*lregularization*): Overfitting is avoided, and favorable features in the resulting SR images are encouraged by regularization loss. Typical regularization methods include adversarial loss to impose realism, total variation regularization to enhance smoothness, and L1 or L2 regularization on the generator's parameters. The regularization loss can be expressed mathematically as follows:

$$lregularization = \lambda_{reg} = \sum_{\theta}\|\theta\|_P \quad (2)$$

Where θ stands for the generator model's parameters, P is the regularization norm (e.g., $P = 1$ for L1 regularization and $P = 2$ for L2 regularization), and *lregularization* is the regularization parameter.

The total loss for the single-image SR is then computed as the weighted sum of the perceptual loss and regularization loss:

lSSR = λperceptual·lperceptual + λregularization ·lregularization

Where λperceptual and λregularization are hyperparameters that control the relative importance of the perceptual and regularization terms, respectively.

3. Results and Discussion

3.1. Dataset

In this work, a new dataset is introduced, consisting exclusively of images captured by the Linear Imaging Self-Scanning (LISS) satellite sensor, as shown in Fig. 1. The LISS sensor provides fine-grained spatial resolutions of 24 m and 5 m, while capturing images across a 140-km orbital sweep. Every 24 days, its operational cycle is repeated. In contrast, the Advanced Wide Field Sensor (AWiFS) offers a slightly poorer spatial resolution of 56 m but spans an even larger orbital sweep of 740 km. AWiFS also features a more frequent revisit cycle of 5 days.

The original LISS images had a pixel resolution of 1956 × 1983. These images were randomly cropped to 86 × 86 to generate training and validation datasets. After that, the cropped images underwent image degradation and were downsampled by a factor of four to produce an LR image set. This LR set was paired with the original image to form LR–HR training pairs. Finally, data augmentation techniques, such as rotation and staggering, were used to augment the dataset.

3.2. Training Process

The proposed architecture was trained in an experimental setup over 40 training epochs with a batch size of 32 and a learning rate of 0.001. Training GAN-type networks presents inherent challenges,

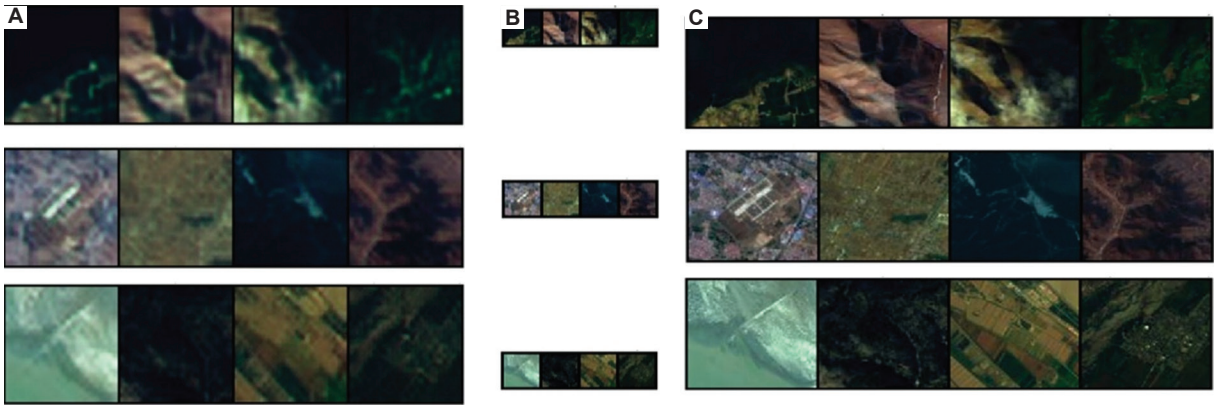

Fig. 1. Original images. (A) Ground truth images. (B) Low-resolution images. (C) High-resolution images

such as instability and mode collapse. To address these issues, a halting standard was introduced. In particular, if a predefined loss threshold of 0.001 was not achieved after 400 epochs, training was terminated. The model loss was monitored after each epoch, and the checkpoint with the lowest loss was preserved. By acting as a benchmark for subsequent epochs, this checkpoint ensured the model maintained optimal performance. The goal of this approach was to reduce instability, prevent premature convergence, and enhance model performance.

Despite some fluctuations, the loss trend exhibited a steadily dropping trajectory, suggesting that the model was improving and converging. Notably, the proposed model showed improved stability and resilience compared to conventional GAN networks, where variations in the loss function could reflect training collapse and intrinsic instability. To produce LR images, each HR image was downscaled by a factor of four. The LR images were 32 × 32 pixels in size, while the HR images were 128 × 128 pixels. Training on smaller image patches enabled the model was able to concentrate on fine-grained local textures, structural components, and object characteristics in remote sensing images. This method facilitated the extraction of crucial information and patterns needed for accurate SR reconstruction. Additionally, using smaller-sized images reduced memory utilization and computational complexity, contributing to a more efficient training process.

3.3. Dataset Splitting Strategy

Three datasets were used, covering varied remote sensing scenarios. Each dataset was split into 70% for training, 15% for validation, and 15% for testing, ensuring the model was evaluated on unseen data to assess generalization performance.

The training details are as follows:

- Epochs: Trained for 40 epochs (with early stopping after 400 if loss <0.001 not achieved)
- Batch size: 32
- Learning rate: 0.001
- Loss monitoring: Checkpoints were saved based on the minimum validation loss to prevent overfitting and mode collapse
- Downscaling: HR images (128 × 128) were downscaled to 32 × 32 for training, reducing memory usage while retaining crucial local features.

In the objective evaluation shown in Table 1, two metrics, including structural similarity index measure (SSIM) and peak signal-to-noise ratio (PSNR), were chosen to quantitatively assess the quality of the super-resolved images.

3.4. Qualitative Analysis

The proposed method was compared to a CNN-based method (referred to as VDSR) and two GAN-based methods (SRGAN and ESRGAN), which integrate perceptual loss through fusion methods to enhance visual quality (Fig. 2). These SR algorithms were all meticulously optimized on the training set for fair comparison. The proposed method primarily relies on a CNN-based approach with a generator network, trained using pixel loss to independently reconstruct HR images from LR images. It comprises both generator and discriminator networks within a GAN model, incorporating perceptual loss for improved visual quality. However, it may lack human perception due to its sole reliance on pixel loss.

The evaluation compared the proposed method against conventional single-image SR algorithms across three test sets. Two different networks (ResNet-50 and modified VGG-19) were used in the deep ResNet–GANs approaches, ensuring a balanced analysis. The proposed method generates realistic and visually appealing results closely resembling natural images, attributed to innovative algorithm design

Table 1. Different objective evaluation metrics

Parameter	Equation	Explanation
Peak signal-to-noise ratio (PSNR)	$PSNR=10\ log_{10}\ (MAX^2/MSE)$ where MAX is the maximum pixel value of the original image data, and mean squared error (MSE) is the average of the squared deviations between the comparable pixel values of the original and super-resolved images.	A key statistic for assessing the quality of super-resolved images in relation to their original counterparts is the PSNR. In remote sensing, the PSNR is a vital measure of the performance of super-resolution techniques in preserving image information and reducing reconstruction mistakes. A higher PSNR value indicates a better fidelity and less distortion between the original and super-resolved images.
Structural similarity index (SSIM)	$SSIM\ (x, y) = (2\mu x\mu y+c1)\ (2\sigma xy+c2)/(\mu x2 + \mu y2+c1)\ (\sigma x2 + \sigma y2+c2)$ where x and y represent the original and super-resolved images, respectively; μx and μy denote the mean intensity of x and y images; σx and σy are the standard deviations of x and y images; σxy represents the covariance of x and y images; and $c1$ and $c2$ are constants added to avoid instability when the denominator approaches zero.	The SSIM metric compares the brightness, contrast, and structure of two images to determine their similarity. It yields a number between 0 and 1, where 0 represents total dissimilarity and 1 represents perfect similarity. As SSIM accounts for aspects of human visual perception, unlike standard metrics, such as MSE or PSNR, it is more appropriate for evaluating image quality changes that impact human perception. This makes it especially pertinent to remote sensing applications.

Fig. 2. Comparison of reconstructed details in distinct small targets (i and ii): (A) original image; (B) images reconstructed using VDSR; (C) images reconstructed using SRGAN; (D) images reconstructed using ESRGAN; and (E) images reconstructed using the proposed algorithm in this study
Abbreviations: ESRGAN: Enhanced super-resolution generative adversarial network; SRGAN: Super-resolution generative adversarial network; VDSR: Very deep super-resolution

techniques that balance visual quality optimization and artifact minimization.

3.5. Quantitative Analysis

In this study, two metrics were employed to quantitatively evaluate the SR results: PSNR and SSIM.

A comparison of the PSNR scores for several SR techniques—VDSR, SRGAN, ESRGAN, and the suggested method—across three different datasets is shown in Table 2.

(i) Interpretation of the PSNR metric: A quantitative indicator of image quality; a greater PSNR value denotes a smaller disparity between the original and reconstructed images

(ii) Superiority of the proposed technique: The PSNR scores obtained by the proposed approach are consistently higher than those obtained by VDSR, SRGAN, and ESRGAN on all three datasets

(iii) Dataset-specific performance: Extensive PSNR values for every dataset demonstrate the enhanced image quality preservation performance of the proposed approach

Table 2. Average PSNR/dB for different algorithms

Dataset	VDSR	SRGAN	ESRGAN	Proposed method
1st dataset	28.345	26.484	32.896	33.256
2nd dataset	29.275	26.784	31.569	32.886
3rd dataset	29.568	29.725	32.451	34.885

Abbreviations: ESRGAN: Enhanced super-resolution generative adversarial network; PSNR: Peak signal-to-noise ratio; SRGAN: Super-resolution generative adversarial network; VDSR: Very deep super-resolution.

(iv) Low difference in the first dataset: The proposed approach produces a PSNR of 33.256 in the first dataset, demonstrating a small difference between the reconstructed and original images, and a higher level of image quality than the other methods
(v) Consistent performance in the second dataset: The proposed method consistently reduces disparities in image reconstruction, as seen by its high PSNR across datasets, with a score of 32.886 in the second dataset
(vi) Highest PSNR in the third dataset: The proposed method outperforms other methods in producing better image quality and fidelity, as evidenced by its PSNR value of 34.885, the highest in the third dataset
(vii) Implication of results: The higher PSNR values of the proposed method across all tested datasets indicate its usefulness and resilience in a range of settings, suggesting that it excels in reducing disparities between reconstructed and real images for improved image quality.

The average SSIM scores for the proposed method, VDSR, SRGAN, ESRGAN, and other SR methods across several datasets are shown in Table 3.

i. The SSIM compares the brightness, contrast, and structure of the reconstructed and original images; higher values within the range of [0,1] indicate better preservation of these features. Out of all the methods tested, the proposed method recorded the highest SSIM scores, demonstrating its superior ability to maintain image properties, including structural similarities, brightness, and contrast, during the SR process.
ii. In the first dataset, for example, the proposed method achieves an SSIM score of 0.862, indicating that it can maintain image properties more effectively than other methods. It also demonstrates outstanding performance in other datasets, as evidenced by the scores as high as 0.833 in the second dataset and 0.899 in the third.
iii. The higher SSIM values of the proposed method across several datasets highlight its extraordinary performance in preserving image features, indicating its potential for producing high-quality SR outcomes with improved visual integrity.

Table 3. Average SSIM for different algorithms

Dataset	VDSR	SRGAN	ESRGAN	Proposed method
1st dataset	0.745	0.771	0.772	0.862
2nd dataset	0.725	0.774	0.702	0.833
3rd dataset	0.812	0.698	0.806	0.899

Abbreviations: ESRGAN: Enhanced super-resolution generative adversarial network; SRGAN: Super-resolution generative adversarial network; SSIM: Structural similarity index measure; VDSR: Very deep super-resolution.

4. Conclusion

In conclusion, this work introduces a novel approach utilizing ResNet-50 in the generator network of the stable SRGAN, alongside a modified VGG-19 network in the discriminator. Through employing bypass branches and shortcuts in ResNet-50, the network effectively learns residuals and mitigates information loss during transmission. This design, combined with the enhanced feature extraction capabilities of the modified VGG-19 network, improves the overall performance of the SR process through preserving image properties and discerning between real and generated images.

The proposed method was compared against conventional single-image SR algorithms, including VDSR, SRGAN, and ESRGAN, across three datasets. The results consistently demonstrate the superiority of the proposed approach, as evidenced by higher SSIM and PSNR scores. For example, achieving an SSIM score of 0.862 in the first dataset signifies superior preservation of image properties compared to other methods. Furthermore, notable PSNR scores, such as 33.256, 32.886, and 34.885 across the datasets, highlight the capability of the proposed approach to maintain image quality and fidelity.

While the proposed approach demonstrates strong performance in terms of quantitative metrics and architectural innovation, there are a few areas that present opportunities for further enhancement. The use of deep networks, such as ResNet-50 and a modified VGG-19, while beneficial for feature extraction and residual learning, may introduce increased computational demands. The evaluation primarily relies on PSNR and SSIM scores; incorporating perceptual or user-centered assessments in future studies could provide a more comprehensive understanding of visual quality. Lastly, as with many GAN-based models, ensuring stable and efficient training across diverse conditions may benefit from further optimization and experimentation.

Future work should focus on enhancing both the efficiency and generalizability of the proposed approach. This includes exploring adaptive or hybrid loss functions—combining perceptual, adversarial, and contextual losses—to improve visual quality and structural fidelity. To support real-world applicability, evaluations should be extended to remote sensing tasks, such as land cover classification and object detection. Additionally, efforts should be made to develop lightweight architectures for resource-constrained environments and to incorporate domain adaptation techniques, enabling robust performance across

varying sensors, regions, and seasonal conditions. These directions aim to refine model effectiveness while broadening its scope and practical relevance.

Acknowledgement

We thank the management and staff of GSSSIETW for providing timely support for completing the work. This research was funded by the "Consolidated University Research for Innovation and Excellence CORE Grant for Women PG College" (DST-CURIE), Department of Science and Technology, Govt. of India, New Delhi, India (Grant no.: DST/CURIE-PG/2022/71).

References

Adarsh, P., Rathi, P., & Kumar, M. (2020). YOLO v3-tiny: Object detection and recognition using one stage improved model. *2020 6th International Conference on Advanced Computing and Communication Systems (ICACCS), Coimbatore, India*. IEEE, United States, p687–694. https://doi.org/10.1109/ICACCS48705.2020.9074315

Bu, L., Zhang, J., Zhang, Z., Yang, Y., & Deng, M. (2024). Deep learning for integrated speckle reduction and super-resolution in multi-temporal SAR. *Remote Sensing*, 16(1), 18. https://doi.org/10.3390/rs16010018

Dileep, R., Jayanth, J., Choodarathnakar, A.L., & Ravikiran, H.K. (2025). Walrus optimization algorithm for panchromatic and multispectral image fusion. *Remote Sensing Applications: Society and Environment*, 38, 101562. https://doi.org/10.1016/j.rsase.2025.101562

Dileep, R., Jayanth, J., Ravikiran, H.K., Vedha, A.S., Shyamala, C., & Yuvaraju, T. (2024). Enhancing image fusion via optimized BAT algorithm in Brovey transform for remote sensing images. In: *2024 International Conference on Knowledge Engineering and Communication Systems (ICKECS)*. p1–6. https://doi.org/10.1109/ICKECS61492.2024.10617149

Dong, R., Zhang, L., & Fu, H. (2022). RRSGAN: Reference-based super-resolution for remote sensing image. *IEEE Transactions on Geoscience and Remote Sensing*, 60, 1–17. https://doi.org/10.1109/TGRS.2020.3046045

Frizza, T., Dansereau, D.G., Seresht, N.M., & Bewley, M. (2022). Semantically accurate super-resolution generative adversarial networks. *Computer Vision and Image Understanding,* 221, 103464. https://doi.org/10.1016/j.cviu.2022.103464

Han, R, Mei, B., Huang, X., Xue, H., Jiang, X., & Yang, S. (2023). Remote sensing image super-resolution adversarial network based on reverse feature fusion and residual feature dilation. *IEEE Access*, 11, 85259–85267. https://doi.org/10.1109/ACCESS.2023.3304050

Haut, J.M., Fernandez-Beltran, R., Paoletti, M.E., Plaza, J., Plaza, A., & Pla, F. (2018). A new deep generative network for unsupervised remote sensing single-image super-resolution. *IEEE Transactions on Geoscience and Remote Sensing,* 56(11), 6792–6810.

He, Z., Li, J., Liu, L., He, D., & Xiao, M. (2022). Multiframe video satellite image super-resolution via attention-based residual learning. *IEEE Transactions on Geoscience and Remote Sensing*, 60, 1–17. https://doi.org/10.1109/TGRS.2021.3072381 https://doi.org/10.1109/TGRS.2018.2843525

Jayanth, J., Ravikiran, H.K., Shyamala, C., & Dileep, R. (2025). Detecting the stages of ragi crop diseases using satellite data in villages of Nanjangud Taluk. In: Singh, S., Sood, V., Srivastav, A.L., & Ampatzidis, Y., editors. *Hyperautomation in Precision Agriculture.* Academic Press, United States, p131–145. https://doi.org/10.1016/B978-0-443-24139-0.00011-7

Jiang, K., Wang, Z., Yi, P., Jiang, J., Xiao, J., & Yao, Y. (2018). Deep distillation recursive network for remote sensing imagery super-resolution. *Remote Sensing*, 10(11), 1700. https://doi.org/10.3390/rs10111700

Khan, B., Mumtaz, A., Zafar, Z., Sedkey, M., Benkhelifa, E., & Moazam, M. (2023). CGA-Net: Channel-wise gated attention network for improved super-resolution in remote sensing imagery. *Machine Vision and Applications*, 34, 128. https://doi.org/10.1007/s00138-023-01477-0

Lei, S., Shi, Z., & Zou, Z. (2017). Super-resolution for remote sensing images via local-global combined network. *IEEE Geoscience and Remote Sensing Letters*, 14(8), 1243–1247. https://doi.org/10.1109/LGRS.2017.2705985

Li, J., Zheng, K., Gao, L., Han, Z., Li, Z., & Chanussot, J. (2025). Enhanced deep image prior for unsupervised hyperspectral image super-resolution. *IEEE Transactions on Geoscience and Remote Sensing*, 63, 1–18. https://doi.org/10.1109/TGRS.2025.3531646

Li, J., Zheng, K., Gao, L., Ni, L., Huang, M., & Chanussot, J. (2024). Model-informed multistage unsupervised network for hyperspectral image super-resolution. *IEEE Transactions on Geoscience and Remote Sensing*, 62, 1–17. https://doi.org/10.1109/TGRS.2024.3391014

Li, J., Zheng, K., Gao, L., Ni, L., Xiao, M., & Chanussot, J. (2023). X-shaped interactive autoencoders with cross-modality mutual learning for unsupervised hyperspectral

image super-resolution. *IEEE Transactions on Geoscience and Remote Sensing*, 61, 1–17. https://doi.org/10.1109/TGRS.2023.3300043

Li, J., Zheng, K., Liu, W., Li, Z., Yu, H., & Ni, L. (2023). Model-guided coarse-to-fine fusion network for unsupervised hyperspectral image super-resolution. *IEEE Geoscience and Remote Sensing Letters*, 20, 1–5. https://doi.org/10.1109/LGRS.2023.3309854

Li, Q., Yang, R., Xiao, F., Bhanu, B., & Zhang, F. (2022). Attention-based anomaly detection in multi-view surveillance videos. *Knowledge-Based Systems,* 252, 109348. https://doi.org/10.1016/j.knosys.2022.109348

Liu, J., Yuan, Z., Pan, Z., Fu, Y., Liu, L., & Lu, B. (2022). Diffusion model with detail complement for super-resolution of remote sensing. *Remote Sensing*, 14(19), 4834. https://doi.org/10.3390/rs14194834

Min, J., Lee, Y., Kim, D., & Yoo, J. (2024). Bridging the domain gap: A simple domain matching method for reference-based image super-resolution in remote sensing. *IEEE Geoscience and Remote Sensing Letters*, 21, 1–5. https://doi.org/10.1109/LGRS.2024.3345678

Pang, B., Zhao, S., & Liu, Y. (2023). The use of a stable super-resolution generative adversarial network (SSRGAN) on remote sensing images. *Remote Sensing*, 15(20), 5064. https://doi.org/10.3390/rs15205064

Ravikiran, H.K., Jayanth, J., Sathisha, M.S., Yogeesh, G.H., & Dileep, R. (2024). Classification of crop across heterogeneous landscape through experienced artificial bee colony. *SN Computer Science*, 5, 428–435. https://doi.org/10.1007/s42979-024-02790-9

Shi, Y., Jiang, C., Liu, C., Li, W., & Wu, Z. (2023). A super-resolution reconstruction network of space target images based on dual regression and deformable convolutional attention mechanism. *Electronics*, 12(13), 2995. https://doi.org/10.3390/electronics12132995

Sui, J., Ma, X., Zhang, X., & Pun, M.O. (2023). GCRDN: Global context-driven residual dense network for remote sensing image superresolution. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 16, 4457–4468. https://doi.org/10.1109/JSTARS.2023.3273081

Tang, E., Wang, L., Wang, Y., Yu, Y., & Zeng, X. (2022). Lightweight frequency-based attention network for image super-resolution. *Journal of Electronic Imaging*, 31(5), 053014. https://doi.org/10.1117/1.JEI.31.5.053014

Veganzones, M.A., Simões, M., Licciardi, G., Yokoya, N., Bioucas-Dias, J.M., & Chanussot, J. (2016). Hyperspectral super-resolution of locally low rank images from complementary multisource data. *IEEE Transactions on Image Processing*, 25(1), 274–288. https://doi.org/10.1109/TIP.2015.2496263

Wang, C., Zhang, X., Yang, W., Li, X., Lu, B., & Wang, J. (2023). MSAGAN: A new super-resolution algorithm for multispectral remote sensing image based on a multiscale attention GAN network. *IEEE Geoscience and Remote Sensing Letters*, 20, 1–5. https://doi.org/10.1109/LGRS.2023.3258965

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*, 2021, 9829706. https://doi.org/10.34133/2021/9829706

Wang, P., Bayram, B., & Sertel, E. (2022). A comprehensive review on deep learning based remote sensing image super-resolution methods. *Earth-Science Reviews*, 232, 104110. https://doi.org/10.1016/j.earscirev.2022.104110

Wang, X., Sun, L., Chehri, A., & Song, Y. (2023). A review of GAN-based super-resolution reconstruction for optical remote sensing images. *Remote Sensing*, 15(20), 5062. https://doi.org/10.3390/rs15205062

Wang, X., Yi, J., Guo, J., Song, Y., Lyu, J., Xu, J., *et al*. (2022). A review of image super-resolution approaches based on deep learning and applications in remote sensing. *Remote Sensing*, 14, 5423. https://doi.org/10.3390/rs1421542

Wang, Y., Lyu, B., Shi, C., & Hu, Y. (2023). Non-parametric simulation of random field samples from incomplete measurements using generative adversarial networks. *Georisk: Assessment and Management of Risk for Engineered Systems and Geohazards*, 18, 60–84. https://doi.org/10.1080/17499518.2023.2222383

Wang, Y., Shao, Z., Lu, T., Wang, J., Cheng, G., Zuo, X., & Dang, C. (2024). Remote sensing pan-sharpening via cross-spectral-spatial fusion network. *IEEE Geoscience and Remote Sensing Letters*, 21, 37844. https://doi.org/10.1109/LGRS.2023.3337844

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

Zhang, T., Chen, H., Chen, S., & Bian, C. (2022). Edge-enhanced efficient network for remote sensing image super-resolution. *International Journal of Remote Sensing*, 43(14), 5324–5348. https://doi.org/10.1080/01431161.2022.2076217

Zhang, Z., Gao, K., Wang, J., Min, L., Ji, S., Ni, C., & Chen, D. (2022). Gradient enhanced dual regression network: Perception-preserving super-resolution for multi-sensor remote sensing imagery. *IEEE Geoscience and Remote Sensing Letters*, 19, 1-5. https://doi.org/10.1109/LGRS.2021.3134798

AUTHOR BIOGRAPHIES

J. Jayanth is a Professor in the Department of Electronics and Communication Engineering, GSSS Institute of Engineering and Technology for Women, Mysuru, and is a distinguished academic with over 15 years of teaching and research experience. He is an active researcher and has completed funded projects and consultancy projects. JJ completed his B.E., M.Tech., and Ph.D. in Electronics and Communication Engineering from Visveswarya Technological University, Belagavi. His research focuses on image processing and remote sensing. As a prolific author, he has published more than 25 research articles in reputed international journals indexed in the SCI and Scopus databases, as well as in prestigious IEEE conferences. His expertise spans artificial intelligence, machine learning, and antenna design and its analysis. As an educator, he has mentored numerous students, integrating theoretical and practical knowledge to prepare them for successful careers.

H. K. Ravikiran was born in Karnataka, India, in April 1987. He received his B.E. degree in Electronics and Communication Engineering from S.J.C. Institute of Technology, Chikballapur, Karnataka, India, in 2008, and his M.Tech. in Digital Electronics and Communication Systems from Malnad College of Engineering, Hassan, Karnataka, India, in 2010. He completed his Ph.D. in Electrical & Electronics Engineering Science from Visvesvaraya Technological University (VTU), Karnataka, India, in 2020. RK is currently serving as a Professor in the Department of Electronics and Communication Engineering at Navkis College of Engineering, Hassan, with 15 years of teaching and research experience. He has made significant contributions to research, publishing 32 papers across Scopus-indexed journals, SCI journals, international journals, and conferences. He has authored two books on the Internet of Things (IoT) and holds three granted patents and a copyright. He is also an active reviewer for prestigious journals and conferences, contributing to the global research community with his expertise and insights. His research interests include image processing, deep learning, and soft computing techniques, where he continues to pioneer innovative methodologies and impactful solutions.

T. Yuvarju is an Assistant Professor in the Department of Electronics and Communication Engineering, GSSS Institute of Engineering and Technology for Women, Mysuru, with over 14 years of teaching and research experience. He completed his B.E. in Electronics and Communication from KVG Sullia and M.Tech. in Digital Electronics and Communication from M.V.J. College of Engineering, Bangalore. He is currently pursuing a Ph.D. in remote sensing from Visveswarya Technological University, Belagavi. With a strong academic background and a passion for research, YT has made significant contributions to the fields of image processing and remote sensing. His research interests include data analytics, machine learning, and artificial intelligence. He has published research articles in renowned journals and IEEE conferences, establishing himself as a prolific researcher in his field. YT is also actively involved in academic and professional development initiatives, participating in international workshops and conferences to disseminate his findings and collaborate with experts worldwide.

R. Dileep is an Assistant Professor in the Department of Electronics and Communication Engineering, GSSS Institute of Engineering and Technology for Women, Mysuru, with over 14 years of teaching and research experience. He completed his B.E. in Electronics and Communication from RNSIT Bangalore and M.Tech. in Digital Electronics and Communication from Malnad College of Engineering, Hassan. He is currently pursuing a Ph.D. in remote sensing from Visveswarya Technological University, Belagavi. With a strong academic background and a passion for research, DR has made significant contributions to the fields of image processing and remote sensing. His research interests include data analytics, machine learning, and artificial intelligence. He has published research articles in renowned SCI-indexed journals and IEEE conferences, establishing himself as a prolific researcher in his field. DR is also actively involved in academic and professional development initiatives, participating in international workshops and conferences to disseminate his findings and collaborate with experts worldwide.