GAN-Based Re-colorization of SAR Grayscale Images

*Note: Sub-titles aren't captured in Xplore and should not be used

Aryesh
School of Computer Engineering
Kalinga Institute of Industrial Technology
Bhubaneswar, India
aryesh1014@gmail.com

Mayank Krishna
School of Computer Engineering
Kalinga Institute of Industrial Technology
Bhubaneswar, India
mayank4451@gmail.com

Ankush Maiti
School of Computer Engineering
Kalinga Institute of Industrial Technology
Bhubaneswar, India
ankushmaiti7@gmail.com

Lipika Mohanty
School of Computer Science and Engineering
Kalinga Institute of Industrial Technology
Bhubaneswar, India
lipika.mohantyfcs@kiit.ac.in

Adyasha Dash
School of Computer Science and Engineering
Kalinga Institute of Industrial Technology
Bhubaneswar, India
adyasha.dashfcs@kiit.ac.in

***Abstract*—The problem addressed in this study is the challenge of re-coloring SAR grayscale images while maintaining their structural integrity and enhancing their visual and analytical utility. To achieve this, a GAN-based model was developed, where a generator synthesizes realistic colorization and a discriminator ensures quality through adversarial learning. The model was trained on paired SAR grayscale and optical images and evaluated using SSIM, PSNR, and qualitative assessment. Final results showed that following approach outperformed conventional methods, generating colorized images that closely resemble real optical images while preserving SAR features. This method successfully bridges the gap between grayscale SAR and interpretable optical imagery, with future improvements focusing on refining color accuracy and enhancing generalization across datasets.**

***Index Terms*—GANs, SAR Image Colorization, cGAN, PatchGAN, U-Net, Remote Sensing, SAR, Translation from image to image, Perceptual Loss, SSIM, PSNR, LPIPS, Sentinel-1, Sentinel-2, Deep Learning, Image Enhancement, Adversarial Learning, Optical Image Reconstruction.**

I. Introduction

A. Background

SAR [1] captures images of the surface of the earth using microwave signals. Unlike optical imaging, which depends on visible or infrared light, SAR operates effectively in all-weather conditions, penetrating cloud cover, fog, and even darkness. This capability makes SAR very effective for applications such as to monitor the environment, response after disaster, defense surveillance, and planning for urban components.

SAR images provide rich structural and textural information by capturing surface roughness, moisture content, and dielectric properties. These properties allow SAR to reveal features that may be obscured in optical images, making it a reliable tool in adverse conditions. However, external factors such as heavy rainfall can slightly degrade SAR image quality, although its robustness remains superior to many other imaging techniques.

B. Motivation

Despite its advantages, SAR imaging presents challenges in terms of visual interpretation. The primary issue is that SAR images are inherently grayscale, lacking the intuitive color information that aids human perception and analysis. This limitation makes it difficult for non-experts to interpret the data and extract meaningful insights. Furthermore, SAR images suffer from speckle noise, [2] a form of multiplicative noise caused by the coherent processing of microwave signals. This noise degrades image clarity and makes object recognition, segmentation, and classification more complex.

Existing methods for SAR image enhancement, such as filtering and denoising techniques, struggle to balance noise reduction and feature preservation. Conventional approaches often fail to restore fine details and structural integrity while enhancing interpretability. This challenge has motivated for more development in advanced DL-based approaches, particularly GANs, to improve SAR image usability through realistic colorization.

C. Contribution

This research introduces a Conditional GAN (cGAN)-based approach for SAR image colorization, aiming to refuce difference between grayscale SAR images & their optical counterparts. The main components of this study are:

- GAN-Based Image Translation: We propose a U-Net-based generator combined with a PatchGAN discriminator to enhance the realism and structural integrity of colorized SAR images.
- Improved Perceptual Quality: By incorporating perceptual loss from a VGG-19 network, the model ensures color consistency and high-level semantic similarity to real optical images.
- Robust Training Strategy: We address training instability issues by employing label smoothing and carefully tuning the learning rate to balance the generator and discriminator.
- Comprehensive Evaluation: Model's performance is validated by PSNR, SSIM, & LPIPS metrics, demonstrating superior image quality compared to conventional colorization techniques.

This research provides a significant step toward improving SAR image interpretability, facilitating better decision-making in remote sensing applications. Future work will focus on further refining color accuracy and enhancing model generalization across diverse SAR datasets.

II. Literature Review

This research provides a significant step toward improving SAR image interpretability, facilitating better decision-making in remote sensing applications. Future work will focus on further refining color accuracy and enhancing model generalization across diverse SAR datasets.

Various approaches have been proposed to improve SAR image colorization by leveraging deep learning and traditional signal processing techniques. Deng et al. [3] utilized scattering characteristics to assign colors to SAR images, helping to distinguish objects based on their dielectric properties. However, this rule-based method is limited in its generalization to complex terrains. With the use of DL, GANs have become a prominent choice for SAR images to optical image translation. Isola et al. [4] cGANs, which later inspired SAR colorization models. In particular, Shen et al. [5] proposed IcGAN4ColSAR, a multispectral cGAN approach tailored for SAR images, demonstrating improvements in spectral fidelity. Ji et al. [6] extended this idea with a multidomain cycle-consistency GAN to refine texture preservation during colorization.

In parallel, Schmitt et al. [7] introduced the Sentinel 1 & 2 dataset, which has become a benchmark for SAR-optical fusion models. They proposed a variational autoencoder-based approach to improve the realism of generated images while retaining SAR-specific properties. Shen et al. [8] further contributed by establishing a benchmarking protocol, highlighting the importance of evaluating SAR colorization models using multiple metrics, including PSNR, SSIM, and perceptual similarity scores. These studies provide the foundation for our work.

III. Methodology

A. Dataset

The dataset used for this study is sourced from the University of Munich and consists of a subset of the Sentinel-1 and Sentinel-2 satellite image pairs [9]. Specifically, 16,000 paired SAR and optical images were selected from the dataset published in ISPRS Annals [10] [11]. These images provide a diverse set of real-world conditions, allowing the model to learn meaningful colorization patterns that generalize across different landscapes and environments.

B. Preprocessing

To prepare the dataset for training, dimensions were reduced to 256 × 256 pixels to standardize the inputs. Further images were normalized to a range of [-1, 1] with mean = 0.5 and a S.D. = 0.5, ensuring stable training dynamics. [12]

C. Model Architecture

The model is on the base of cGAN, designed specifically for SAR image colorization. It consists of:

1) Generator: U-Net Architecture: The generator is structured as a UNET. This design helps preserve fine details and structural information while generating realistic colorized SAR images. The U-Net captures both low-level and high-level features, facilitating a better understanding of SAR image patterns.

- *Encoder (Downsampling Path)*: Uses DBlocks (Downsampling Blocks), which consist of Convlayers, BN, and LReLU(2/10). It starts with a single-channel SAR input and progressively increases the number of feature maps, extracting hierarchical features at multiple levels. The encoder layers use kernels of size 4x4 with stride 2 or 1 and padding 1, ensuring the geometric dimensions are reduced while retaining crucial information. These operations allow the model to effectively capture texture, edges, and object boundaries, which are essential for accurate colorization.
- *Bottleneck*: The deepest layers of the network use 512 and 1024 feature maps, ensembles high-level parameters of the image. Layers act as the most abstract representation of the input, learning complex structures and contextual relationships present in SAR images. By transforming the learned features into a compressed representation, the bottleneck ensures that the decoder can reconstruct a meaningful and realistic colorized image with well-defined geometric characteristics.
- *Decoder (Upsampling Path)*: Uses UBlocks (Upsampling Blocks), consisting of Transposed Convolutions, BN, RELU . All decoder layer upsamples feature maps to reconstruct the image with added color channels, progressively refining the details as the resolution increases. Skip connections - encoder to the corresponding decoder

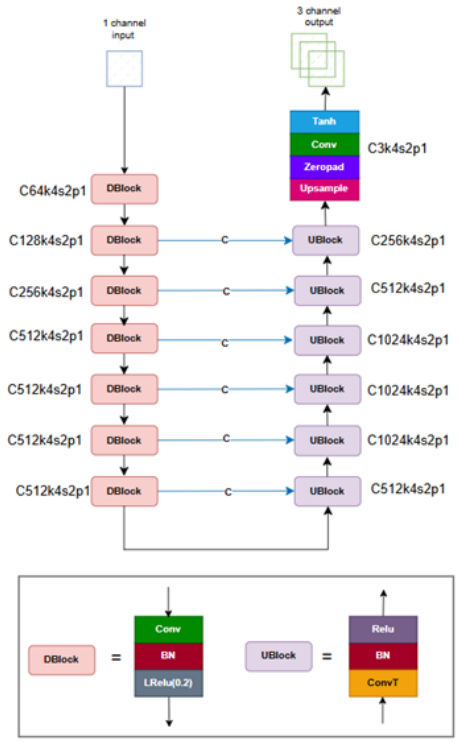

Fig. 1. Generator Architecture

layers retain geometric details & prevent information loss, ensuring the reconstructed image maintains fine details from the original SAR input. This skip-connection mechanism significantly enhances the quality of the generated colorized image, as it combines both high-level semantic understanding and low-level geometric accuracy.

- *Final Output*: A 3-channel colorized image is generated, processed through tanh activation, ensuring the output values are in a normalized range. The last convolution layer (C3k4s2p1) refines the output before passing through Tanh activation, which ensures the pixel intensity values remain within a bounded range, avoiding extreme values that could distort the colorization. This final step guarantees that the output image is not only visually appealing but also consistent in terms of brightness and contrast, closely resembling real-world optical images.

Figure 1 illustrates the U-Net-based generator used in our cGAN model. The encoder extracts hierarchical features from the SAR grayscale image, while colorized image is reconstructed by decoder with skip connections to preserve geometric details. Figure 2 compares a grayscale SAR image (left) with its colorized counterpart (right) generated by our model, demonstrating how well the model translates grayscale information into realistic colorization. This comparison highlights its ability to recover very minute details and complex structures, showcasing its effectiveness in enhancing SAR imagery for better interpretability and visualization.

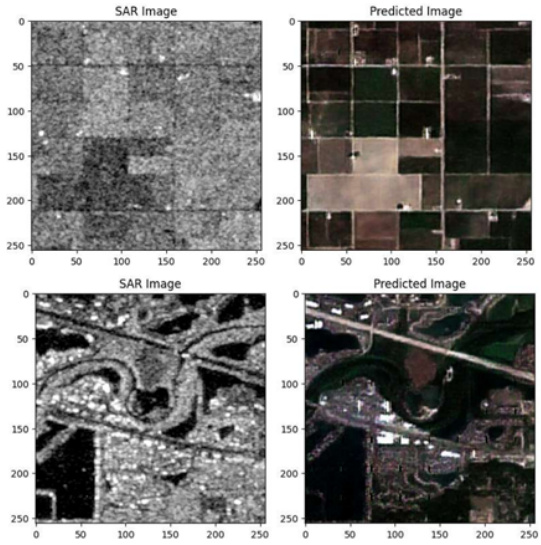

Fig. 2. Sample input image (left) and corresponding generated image (right)

D. Discriminator: PatchGAN Architecture

The discriminator follows a PatchGAN-based approach, which assesses image realism at a local patch level rather than the entire image. This ensures that the generated images maintain local consistency, sharpness, and fine-grained details, making them more visually convincing. By focusing on smaller patches rather than the full image, the discriminator learns to distinguish subtle artifacts more effectively, thereby enhancing adversarial training.

- *Architecture*: The discriminator processes both real and generated images as input. It consists of DNBlock (Downsampling Normalization Block) and multiple DBlocks (Downsampling Blocks). DNBlock includes a convolution layer, batch normalization (BN), and Leaky ReLU (LReLU), ensuring stable training and effective feature extraction. DBlocks continue downsampling through convolution layers with progressively increasing feature maps. The convolution layers follow a structured progression in feature maps ($64 \rightarrow 128 \rightarrow 256 \rightarrow 512$) while reducing the spatial resolution, enabling hierarchical feature extraction. This design allows the model to focus on different levels of abstraction, improving its ability to detect inconsistencies in generated images.
- *Final Layer*: The final layer applies a 4×4 convolution, predicting whether each patch (30×30) corresponds to a real or fake image. Instead of classifying the entire image, it classifies multiple patches independently, improving robustness against small-scale inconsistencies. By using a patch-based approach, the discriminator enforces fine-grained realism, making the generator produce more natural textures and details. Additionally, this method helps mitigate mode collapse by encouraging diversity in generated patterns across different image regions.

This combination of U-Net generator and PatchGAN dis-

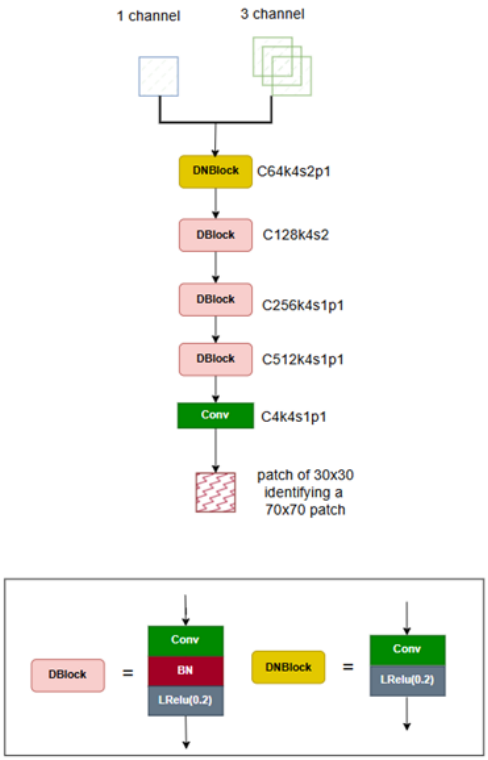

Fig. 3. Discriminator Architecture

criminator makes the cGAN model well-suited for SAR image colorization by ensuring high-quality and locally coherent outputs. Fugure 3 represents the PatchGAN-based discriminator, which evaluates small patches of an image to determine whether they are from optical images or generated by the model, ensuring localized realism in the generated images

E. Loss Functions

To ensure high-quality and realistic colorization, the model employs a many loss functions.

1) BCE: The adversarial loss is computed as a BCE difference between the discriminator's o/p for the images that were generated & a target label corresponding to a "valid" patch. Given that our discriminator learns to be strong, we use a label smoothing approach, assigning the valid patch a value of 0.9 instead of the standard 1.0. The loss is expressed as:

$$L_{adv} = \mathbb{E}_x[\log D(G(x))] + \mathbb{E}_y[\log(1 - D(y))] \quad (1)$$

where, x represents generated images, D denotes the distribution of generated samples, and $y = 0.9$ is the valid patch label.

2) Perceptual Loss: Inspired by Johnson et al. (2016), [13] this loss is computed using a VGG-19 network pretrained on ImageNet. The Block4 Conv4 features are extracted from both the type of images to compute high-level perceptual similarity:

$$L_{perc} = \sum_l \lambda_l \|\phi_l(G(x)) - \phi_l(y)\|_2^2 \quad (2)$$

where, $G(x)$ is the generated image, y is the ground truth optical image, ϕ_l is a feature map from VGG-19, and λ_l is a weighting factor for different layers.

3) Mean Absolute Error (MAE) [14]: Direct pixel-wise comparison between the images to minimize overall reconstruction error:

$$L_{mae} = \frac{1}{N} \sum_{i=1}^{N} |G(x)_i - y_i| \quad (3)$$

where, $G(x)$ is the generated image, y is the ground truth image, and N is the total number of pixels.

The total generator loss is formulated using (1), (2), and (3):

$$L_G = L_{adv} + 100 \times L_{perc} + L_{mae} \quad (4)$$

4) Discriminator Loss: It distinguishes real from generated images. Its loss is two binary cross entropy losses:

$$L_D = L_{BCE}(D(y), 1) + L_{BCE}(D(G(x)), 0) \quad (5)$$

where L_{BCE} is the binary cross-entropy loss for real images labeled with the valid patch (0.9) and L_{BCE} is binary cross entropy of generated images having label with the fake patch (0.1).

IV. Optimization and Training

Training was with the Adam optimizer, ensuring stable and efficient learning. The generator was optimized the learning rate of $2e^{-4}$, while discriminator used Adam (learning rate = $1e^{-5}$). The lower learning rate for the discriminator was intentionally set to slow down its learning process, also allowing the generator to catch up and produce better quality images.

The training was conducted for 100 epochs, batch size = 32 on an NVIDIA RTX 2060 GPU, ensuring efficient computation and convergence.

V. Evaluation Metrics

Performance was evaluated by three evaluation metrics to ensure high-quality image generation.

1) PSNR: Talks about the overall quality of the image that is reconstructed.

$$PSNR = 10 \log_{10} \left(\frac{L^2}{MSE} \right) \quad (6)$$

where, L is maximum pixel intensity and MSE is mean squared error.

2) SSIM [15]: Tells about the structural consistency between generated and target images.

$$SSIM(x, y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \quad (7)$$

where, μ_x, μ_y are mean intensities of images x and y, σ_x, σ_y are standard deviations of x and y, and C_1, C_2 are small constants to avoid division by zero.

3) LPIPS [16]: Quantifies perceptual differences between images.

$$LPIPS(x, y) = \sum_{l} \lambda_l \|\phi_l(x) - \phi_l(y)\|_2^2 \tag{8}$$

where $\phi_l(x)$ and $\phi_l(y)$ are feature maps from layer l of a CNN and ω_l is a learned weight.

By incorporating these techniques, the proposed model effectively bridges the gap between SAR grayscale images and visually interpretable optical images, ensuring both color accuracy and structural integrity.

VI. Results & Discussion

The proposed GAN-based model effectively colorized SAR grayscale images, demonstrating significant improvements over existing approaches. The results for the same are as follows : PSNR, SSIM, and LPIPS, providing a quantitative measurement of both the perceptual & structural quality for generated images. Table I summarizes the quantitative performance metrics (PSNR, SSIM, and LPIPS) used to evaluate the generated colorized images. Higher PSNR and SSIM, along with lower LPIPS, indicate better reconstruction and perceptual similarity to optical images. Additionally, qualitative visual comparisons further ensure effectiveness of the proposed method, achieving enhanced preservation, reduced artifacts, and improved color accuracy in the reconstructed images.

TABLE I
Performance of the model

Metric	Performance
PSNR	21.31 dB
SSIM	0.57
LPIPS	0.3119

These results indicate that our model produces images with higher perceptual quality and greater structural accuracy, confirming the effectiveness of adversarial learning and perceptual loss. Figure 4 visually compares the results produced by our model with the ground truth optical image, showing the level of similarity achieved in colorization.

A. Effect of Perceptual Loss

The performance comparison of cGAN models with and without perceptual loss reveals significant improvements across key image quality metrics. The model incorporating perceptual loss achieves a PSNR of 21.31, notably higher than 16.35 for the model without perceptual loss. This indicates better reconstruction quality and reduced distortion. Similarly, the SSIM score is 0.570 for the perceptual loss model, compared to 0.416 for the model trained without it, demonstrating better structural similarity to the ground truth images. Moreover, the model with perceptual loss records 0.3119 LPIPS, while the perceptual loss-free model exhibits a higher LPIPS value of 0.3742, indicating that the latter loss is unfavorable. These results highlight the critical role

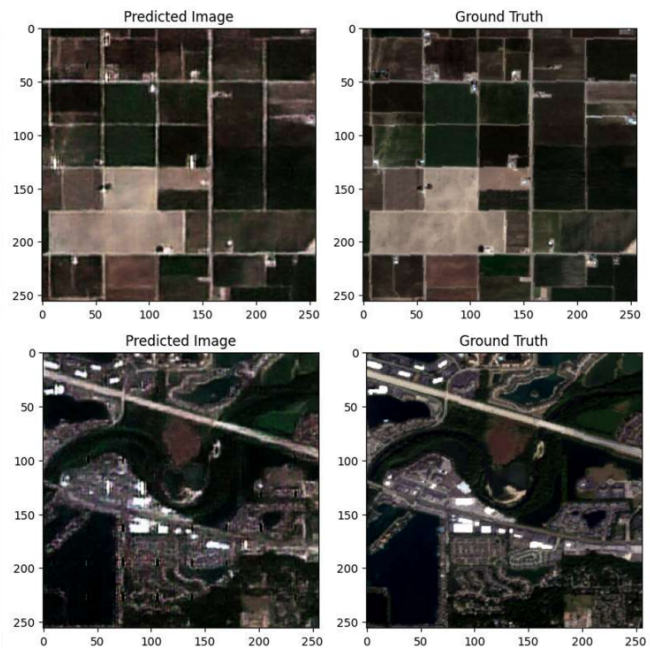

Fig. 4. Sample predicted image (left) and the corresponding ground truth (right).

of perceptual loss in enhancing the realism and structural accuracy of the generated images. Figure 5 shows sample input grayscale image and corresponding generated image and ground truth image.

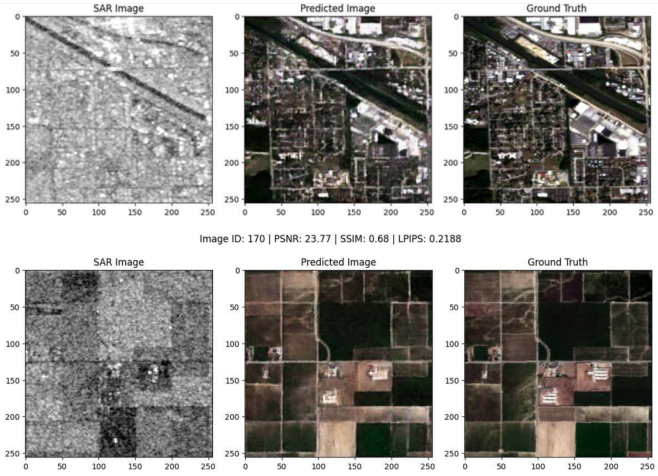

Fig. 5. Sample input SAR grayscale image (left), sample predicted image (middle), sample ground truth (right).

Figure 6 shows a comparison graph of the performance of the model with perpetual loss against without perpetual loss.

Our training strategy effectively optimized the adversarial framework, leading to high-quality generated images. Initially, the discriminator was too strong, preventing the generator from learning effectively. However, by reducing the learning rate and using smoothed labels (0.9 for real images and 0.1 for fake images), we weakened the discriminator slightly, allowing

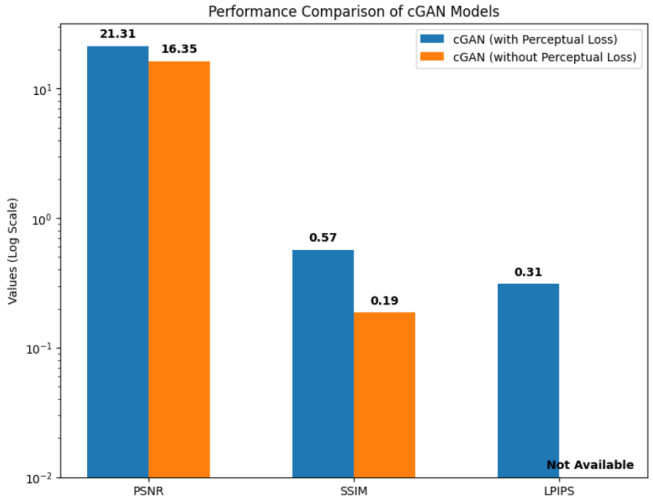

Fig. 6. Comparison plot between cGAN model with perceptual loss and without perceptual loss.

the generator to compete more effectively. This balance led to optimized training dynamics, ultimately improving the generated image quality.

Overall, the model demonstrated strong colorization performance, with perceptual loss driving color accuracy, geometric loss improving smoothness at the cost of edge sharpness, and adversarial loss contributing to realism and overall fidelity. Future improvements could involve refining the edge-preserving mechanism and adapting a more robust loss function for the discriminator to further enhance the quality of colorized SAR images.

VII. Conclusion

This study introduced a Conditional GAN-based approach for SAR grayscale image re-colorization, addressing a key challenge in remote sensing. By integrating adversarial learning, perceptual loss, and mean absolute error loss, the model effectively synthesized colorized SAR images while preserving structural and textural integrity.

Experimental results demonstrated competitive performance across PSNR, SSIM, and LPIPS metrics, with perceptual loss significantly enhancing visual fidelity. Training challenges, such as discriminator dominance, were mitigated through learning rate adjustments and label smoothing, leading to more stable adversarial training and improved image generation quality.

While the results are promising, further improvements in edge preservation, robustness across diverse SAR datasets, and the incorporation of refined loss functions and transformer-based architectures could enhance performance.

Overall, this research underscores the potential of deep learning-driven SAR re-colorization for enhanced image interpretation, benefiting applications in environmental monitoring, disaster response, and land use classification. As SAR images will continue to expand, AI-driven methodologies are anticipated to play a vital role in remote sensing and geogeometric analysis.

References

[1] A. Caglayan, N. Imamoglu, and T. Kouyama, "Sar-w-mixmae: Sar foundation model training using backscatter power weighting," *arXiv preprint arXiv:2503.01181*, 2025.

[2] P. Singh and R. Shree, "Analysis and effects of speckle noise in sar images," in *2016 2nd International Conference on Advances in Computing, Communication, & Automation (ICACCA)(Fall)*. IEEE, 2016, pp. 1–5.

[3] Q. Deng, Y. Chen, W. Zhang, and J. Yang, "Colorization for polarimetric sar image based on scattering mechanisms," in *2008 Congress on Image and Signal Processing*, vol. 4. IEEE, 2008, pp. 697–701.

[4] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 1125–1134.

[5] K. Shen, G. Vivone, S. Lolli, M. Schmitt, X. Yang, and J. Chanussot, "Icgan4colsar: A novel multispectral conditional generative adversarial network approach for sar image colorization," *IEEE Geoscience and Remote Sensing Letters*, 2025.

[6] G. Ji, Z. Wang, L. Zhou, Y. Xia, S. Zhong, and S. Gong, "Sar image colorization using multidomain cycle-consistency generative adversarial network," *IEEE Geoscience and Remote Sensing Letters*, vol. 18, no. 2, pp. 296–300, 2020.

[7] M. Schmitt, L. H. Hughes, and X. X. Zhu, "The sen1-2 dataset for deep learning in sar-optical data fusion," *arXiv preprint arXiv:1807.01569*, 2018.

[8] K. Shen, G. Vivone, X. Yang, S. Lolli, and M. Schmitt, "A benchmarking protocol for sar colorization: from regression to deep learning approaches," *Neural Networks*, vol. 169, pp. 698–712, 2024.

[9] M. Schmitt, "SEN1-2," 2018, accessed on Feb 20, 2025. [Online]. Available: https://mediatum.ub.tum.de/1436631

[10] P. Tiwari, "Sentinel-1&2 image pairs (sar & optical)," March 2021, accessed on Feb 20, 2025. [Online]. Available: https://www.kaggle.com/datasets/requiemonk/sentinel12-image-pairs-segregated-by-terrain

[11] M. Schmitt, L. Hughes, and X. Zhu, "The sen1-2 dataset for deep learning in sar-optical data fusion. arxiv 2018," *arXiv preprint arXiv:1807.01569*.

[12] T. Salimans, I. Goodfellow, W. Zaremba, V. Cheung, A. Radford, and X. Chen, "Improved techniques for training gans," *Advances in neural information processing systems*, vol. 29, 2016.

[13] J. Johnson, A. Alahi, and L. Fei-Fei, "Perceptual losses for real-time style transfer and super-resolution," in *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part II 14*. Springer, 2016, pp. 694–711.

[14] S. Kaji and S. Kida, "Overview of image-to-image translation by use of deep neural networks: denoising, super-resolution, modality conversion, and reconstruction in medical imaging," *Radiological physics and technology*, vol. 12, no. 3, pp. 235–248, 2019.

[15] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE transactions on image processing*, vol. 13, no. 4, pp. 600–612, 2004.

[16] R. Zhang, P. Isola, A. A. Efros, E. Shechtman, and O. Wang, "The unreasonable effectiveness of deep features as a perceptual metric," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 586–595.