From Noise To Insight: A Dual-Stage Deep Learning Approach For SAR Image Denoising And Colourisa- Tion

Arunima Muralitharan
Department of Computer Science and Engineering
Sri Venkateswara College of Engineer-ing
Chennai, India

Kartheesan Senthilkumar *Department of Computer Science and Engineering*
Sri Venkateswara College of Engineer-ing
Chennai, India

Muralidharan S
Department of Computer Science and Engineering
Sri Venkateswara College of Engineer-ing
Chennai, India

***Abstract*— Synthetic Aperture Radar (SAR) provides an invaluable remote sensing capability, operating independently of weather conditions and daylight. However, the usage of SAR imagery is constrained by two inherent characteristics: the presence of multiplicative speckle noise, which degrades image quality and obscures details, and its monochromatic nature, which complicates interpretation by humans and limits its direct use with standard computer vision algorithms. To address these issues, this paper introduces a sequential two-stage deep learning framework designed to holistically enhance SAR imagery. The first stage employs a Convolutional Autoencoder (CAE), specifically designed for speckle reduction. This network learns a robust representation of the underlying scene structure, filtering the granular noise. The second stage uses a Pix2Pix architecture, a conditional GAN (cGAN) framework tailored for paired image-to-image translation tasks, to perform SAR image colourisation. The CAE demonstrates strong denoising performance, achieving a Peak Signal-to-Noise Ratio (PSNR) of up to 32 dB and a Structural Similarity Index Measure (SSIM) of up to 0.88. The subsequent colourisation stage generates visually plausible images, validated by a Fréchet Inception Distance (FID) of 72.3 and a Learned Perceptual Image Patch Similarity (LPIPS) of 0.28. This approach, by decoupling the complex tasks of denoising and colourisation, offers a robust and effective solution, significantly improving the interpretability and analytical value of SAR data for a wide range of geoscience and surveillance applications.**

Keywords—Synthetic Aperture Radar (SAR), Speckle Reduction, Image Colourisation, Deep Learning

I. Introduction

Synthetic Aperture Radar (SAR) has become foundational in modern remote sensing, offering unique imaging capabilities that overcome the limitations of traditional optical sensors. Unlike passive optical systems, SAR actively illuminates the Earth's surface with microwave pulses and records the backscattered signals, enabling high-resolution image acquisition regardless of daylight or weather conditions. By synthesising a large virtual aperture from the motion of the platform, SAR achieves fine spatial resolution that would otherwise require physically impractical antenna sizes.
These characteristics make it highly useful in a wide range of applications, including environmental monitoring, disaster management, geology, agriculture, maritime surveillance, and defence operations [7].

Despite these advantages, SAR imagery presents two major challenges that limit its usefulness. The first arises from the presence of speckle noise, a derivative of the coherent imaging process. Showing as a granular pattern caused by random interference of backscattered radar waves, speckle significantly degrades image quality and complicates both human interpretation and automated analysis. Its multiplicative nature makes it particularly resistant to conventional denoising methods designed for additive noise [9]. Recent work has explored deep learning–based approaches for despeckling, including convolutional autoencoders [17], which demonstrate strong potential for improving SAR image quality. The second issue is the monochromatic format of SAR imagery. With pixel intensity reflecting only the strength of radar backscatter, SAR images lack the spectral richness of optical data. This absence of colour not only reduces their interpretability for human analysts but also restricts compatibility with the vast majority of computer vision models, which are typically optimised for RGB inputs [1].

To address these issues, this work proposes a sequential two-stage deep learning framework that holistically enhances SAR imagery. The first stage employs a Convolutional Autoencoder (CAE) to reduce speckle noise by learning a compact representation of structural information and reconstructing a clean image [2]. The second stage introduces a pix2pix conditional Generative Adversarial Network (cGAN) to colourise the despeckled images, translating them into perceptually plausible RGB representations [3].
This decoupled approach ensures that the GAN operates on stable, noise-free inputs, allowing it to focus solely on the semantic task of assigning meaningful colour. By separating the tasks of denoising and colourisation, the pipeline improves robustness, training efficiency, and output quality.

The contributions of this work are threefold. First, it introduces an end-to-end deep learning framework that systematically enhances SAR imagery through despeckling, followed by colourisation. Second, it demonstrates the effectiveness of this approach through rigorous evaluation, showing that the CAE achieves high fidelity in speckle reduction while the pix2pix GAN produces realistic and visually coherent colourised outputs. Third, it presents a comprehensive performance analysis that integrates both traditional metrics, such as PSNR and SSIM, with modern perceptual measures, including FID and LPIPS. Together, these contributions highlight a pathway to making SAR data more interpretable and actionable for a wide range of geoscience, surveillance, and monitoring applications [5].

II. Literature Review

A. Evolution of SAR Image Despeckling

The reduction of speckle noise has been a central research theme in SAR image processing, with methodologies evolving from classical filtering techniques to advanced deep learning architectures. Early approaches focused on classical filtering methods applied in the spatial domain. Filters such as the Lee, Frost, and Kuan filters operate by analysing local statistical properties within a moving window to estimate the noise-free pixel value. While computationally simple, these methods often suffer from a significant drawback where they tend to blur sharp edges, smooth out fine textures, and consequently lose important image details, representing a trade-off between noise suppression and feature preservation [1].

To overcome these issues, advanced techniques were developed. Transform domain methods, such as those based on the wavelet transform, attempt to separate the signal and noise in the frequency domain [2]. Concurrently, non-local methods emerged, based on the idea of using image self-similarity. The Non-Local Means (NLM) algorithm and its highly successful SAR-specific variant, SAR-BM3D, work by averaging pixels from distant but structurally similar patches across the image. These methods showed better performance in preserving details compared to local filters, but are often associated with high computational complexity [3].

The advent of deep learning has revolutionised SAR despeckling, offering a more powerful, data-driven approach that can learn complex image priors automatically. A variety of neural network architectures have been explored for this task. Directly relevant to the first stage of our proposed framework is the use of Convolutional Denoising Autoencoders (C-DAE), trained to reconstruct clean images from noisy inputs [4]. Other prominent deep learning paradigms include self-supervised approaches, which use the statistical properties of SAR data to train a network without requiring perfectly clean ground-truth images, a significant advantage given that such ground truth is often unavailable [5].

Generative Adversarial Networks (GANs) have also been applied to image restoration, with models trained to generate realistic denoised images that are indistinguishable from clean ones [6]. More recently, the state-of-the-art has been advanced by Denoising Diffusion Probabilistic Models (DDPMs), a class of powerful generative models adapted for despeckling, which have shown remarkable performance in generating high-fidelity results [7]. Furthermore, prior-driven networks have been explored, which explicitly incorporate physical or statistical models of SAR imaging into the network architecture to guide the learning process and improve structure preservation [8].

B. SAR Image Colourisation and SAR-to-Optical Translation

The task of converting monochromatic SAR images into a visually intuitive colour format is most effectively framed as an image-to-image translation problem. The goal is to learn a mapping function that transforms an image from the source domain (SAR) to a target domain (optical RGB). Generative Adversarial Networks have become the go-to tool for this challenging task [9].

Conditional GANs (cGANs) are particularly well-suited for this problem when paired datasets, containing spatially co-registered SAR and optical images of the same scene, are available. The pix2pix model, a foundational cGAN architecture, has been widely adapted for this purpose [10]. Examples have demonstrated the feasibility of using cGANs and other CNN-based architectures to generate high-quality visible images from SAR inputs, showing the potential of learning this complex cross-modal mapping [11].

In scenarios where perfectly aligned image pairs are difficult or impossible to obtain, Cycle-Consistent GANs (CycleGANs) provide an elegant solution for unpaired image-to-image translation. By enforcing a cycle-consistency loss, ensuring that an image translated from domain A to B and back to A recovers the original, models have been able to learn the translation function without direct one-to-one supervision [12], [13].

Recent research has also focused on developing specialised network architectures tailored to the unique characteristics of SAR data. The Sar2color model is a notable example, integrating modules within its GAN architecture to handle speckle noise and geometric distortions during the translation process, aiming to improve the textural and colourimetric accuracy of the generated optical images [14]. The progress in this field is further supported by comprehensive surveys on image colourisation [15] and the creation of large-scale, publicly available benchmark datasets that facilitate the training and evaluation of these data-hungry models [16].

Most research efforts either focus exclusively on the problem of despeckling or on the problem of SAR-to-optical translation. While some translation models, such as Sar2-color, acknowledge speckle as a confounding factor and attempt to mitigate it implicitly within a monolithic architecture, there is a notable gap in the literature regarding the explicit design and validation of a sequential, two-stage pipeline. The approach presented in this paper directly addresses this gap. It posits that by decoupling the two tasks and employing specialised, state-of-the-art networks for each, a more robust and effective overall enhancement can be achieved. Instead of burdening a single complex network with the entangled objectives of denoising and colourisation, this work modularises the problem, allowing each network to perform its designated function optimally.

III. Methodology

The proposed framework is a cascaded system composed of two distinct deep learning models, each tailored for a specific stage of the SAR image enhancement process. This section provides a detailed technical exposition of the system's architecture and training objectives. The input to the system is a noisy, single-channel (grayscale) SAR image, which is first processed by Stage 1, the Denoising Convolutional Autoencoder (CAE). The output of this stage is a denoised grayscale SAR image. This clean image then serves as the input to Stage 2, the Colourisation pix2pix GAN. The GAN performs an image-to-image translation, producing the final output, which is a colourised, three-channel (RGB) image.

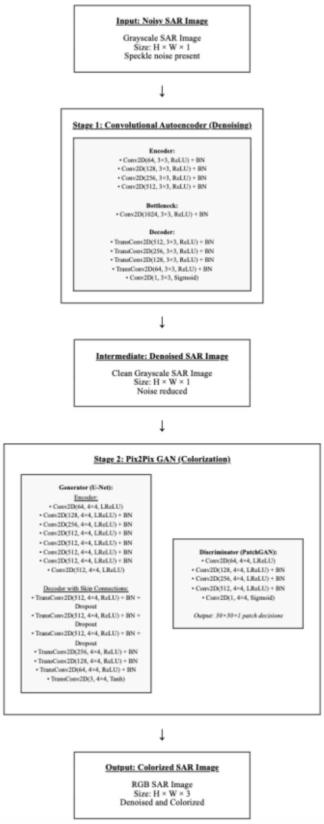


Fig. 1. Architectural overview of the sequential denoising and colourisation pipeline

A. *Training Data and Preprocessing*

The models were trained and evaluated using a combination of publicly available and synthetic datasets to ensure robustness and generalizability. For the denoising stage, the Virtual SAR dataset was utilised, which provides synthetic pairs of clean images and their corresponding versions with artificially added speckle noise [13]. For the colourisation stage, which requires paired SAR and optical data, datasets such as the GRSS SAR/PolSAR DATABASE and the TenGeoP-SARwv dataset, containing over 30,000 Sentinel-1 images, were used [20], [11]. All images across the datasets were subjected to a uniform preprocessing pipeline. Each image was resized to a standard dimension of 256×256 pixels. For Stage 1 (denoising), the pixel values of the grayscale images were normalised to a floating-point range. For Stage 2 (colourisation), the pixel values for both the denoised grayscale input and the target RGB optical images were normalised to the range [-1, 1], a standard practice that aligns with the use of a Tanh activation function in the generator's final layer [1], [3].

B. *Speckle Noise Reduction via Convolutional Denoising Autoencoder (CAE)*

The first stage of the framework is dedicated to reducing the speckle noise common in SAR imagery. A Convolutional Autoencoder is chosen for this task due to its proven effectiveness in learning compressed representations for image reconstruction and denoising [14]. The fundamental principle is that the encoder learns to map the input image to a lower-dimensional latent space that captures the important, structural information of the scene, while the decoder learns to reconstruct the image from this purified latent code, effectively filtering out the noise [5].

The CAE architecture follows a symmetric encoder-decoder structure designed for effective feature extraction and image reconstruction.

Encoder: The encoder consists of a deep stack of Conv2D layers that progressively reduce the spatial dimensions of the input while increasing the feature map depth. The filter count increases hierarchically (e.g., 64, 128, 256, 512, up to 1024), allowing the network to capture features at multiple scales. To stabilise training and accelerate convergence, each convolutional layer is followed by a Batch Normalisation (BN) layer and a ReLU (Rectified Linear Unit) activation function, which introduces non-linearity [14].

Decoder: The decoder mirrors the encoder's structure, employing a symmetric stack of TransposedConv2D layers to perform upsampling and gradually restore the feature maps to the original input dimensions. The number of filters decreases in a mirrored fashion. Similar to the encoder, each TransposedConv2D layer is followed by a BN layer and a ReLU [17] activation.

Output Layer: The final layer of the decoder is a Conv2D layer with a single filter and a Sigmoid activation function. The sigmoid function is crucial as it squashes the output pixel values to the normalised range of, matching the preprocessed ground-truth images [14].

The CAE is trained in a supervised manner, using pairs of noisy SAR images and their corresponding clean, speckle-free ground-truth counterparts. The network's objective is to minimise the error between its output, the denoised image and the clean ground truth [17]. This is achieved by using the Mean Squared Error (MSE) as the loss function. The MSE is defined as:

$$\mathrm{MSE} = \frac{1}{N}\sum_{i=1}^{N}\left(y_i - \hat{y}_i\right)^2 \qquad (1)$$

where y_i is the value of the i-th pixel in the ground-truth image, $\hat{y}_i$ is the value of the corresponding pixel in the CAE's output, and N is the total number of pixels. MSE penalises large errors more heavily and is a standard choice for image reconstruction tasks where pixel-wise fidelity is the primary goal.

C. *Image Colourisation using a Conditional GAN (pix2pix)*

The second stage addresses the challenge of interpretability by translating the denoised grayscale SAR image into a photorealistic colour image. The pix2pix framework is an ideal choice for this task, as it is a conditional GAN specifically designed for paired image-to-image translation problems [1]. It learns a direct mapping from a source image domain (denoised SAR) to a target image domain (optical RGB), conditioned on the input image to ensure structural consistency. The architecture consists of two competing networks, a Generator and a Discriminator [12].

The generator's role is to produce the colourised image. Its architecture is based on a **U-Net**, an encoder-decoder structure enhanced with skip connections, which has proven highly effective for image translation tasks where fine-grained detail must be preserved [3],[10].

Encoder-Decoder Path: The generator follows a classic U-Net structure. The encoder path is composed of a series of

Conv2D layers that progressively downsample the input image, increasing the number of feature filters at each step (e.g., 64, 128, 256, 512) to learn contextual features at varying levels of abstraction. Each convolutional layer is followed by a BN layer and a LeakyReLU activation function. The decoder path symmetrically upsamples the feature maps using TransposedConv2D layers, aiming to reconstruct the full-resolution colour image [1].

Skip Connections: The defining feature of the U-Net architecture is the presence of skip connections that link layers in the encoder directly to their corresponding layers in the decoder. These connections create a shortcut for information flow, allowing low-level spatial information (such as precise edges and textures) captured in the early encoder layers to be directly available to the decoder during image reconstruction. This is critical for generating sharp, detailed outputs.

Regularisation and Output: To prevent overfitting, Dropout layers are strategically placed in the deeper layers of the U-Net. The final layer of the generator uses a Tanh activation function, which maps the output pixel values to the range [-1, 1], consistent with the preprocessed target images.

The discriminator's role is to distinguish between real optical images and the fake colourised images produced by the generator. This framework employs a **PatchGAN** discriminator, which classifies small, overlapping patches of the input image as real or fake. Its output is an N×N feature map, where each element corresponds to the discriminator's verdict on a specific patch. This architecture encourages the generator to focus on producing realistic high-frequency details and textures across the entire image by penalising unrealistic structures at a local level.

The generator (G) and discriminator (D) are trained simultaneously in an adversarial minimax game.

- Adversarial Loss: The framework uses a conditional adversarial loss. The discriminator is trained to correctly classify real image pairs (denoised SAR input, real optical target) and fake image pairs (denoised SAR input, generated colour image). The generator is trained to produce outputs that fool the discriminator. The objective can be expressed as:

$$\min_G \max_D V(D,G) = \mathbb{E}_{\mathbf{x}\sim p_{\text{data}}(\mathbf{x})}[\log D(\mathbf{x})] + \mathbb{E}_{\mathbf{z}\sim p_{\mathbf{z}}(\mathbf{z})}[\log(1 - D(G(\mathbf{z})))] \quad (2)$$

where x is the input denoised SAR image and y is the target optical image.

- L1 Loss: To ensure the generated output is structurally consistent with the input, an L1 loss (Mean Absolute Error) is added to the generator's objective function. This loss measures the pixel-wise absolute difference between the generated image and the ground-truth target image:

$$L_1 = \frac{1}{n}\sum_{i=1}^{n} |y_i - \hat{y}_i| \quad (3)$$

The L1 loss encourages less blurring compared to the L2 loss (MSE) and helps the generator produce outputs that are a plausible translation of the source.

- Combined Loss: The final objective for the generator is a weighted combination of the adversarial loss and the L1 loss, balancing the need for realism with the need for structural fidelity:

$$\min_G \max_D V(D,G) = \mathbb{E}_{x\sim p_{data}(x)}[\log D(x)] + \mathbb{E}_{z\sim p_z(z)}[\log(1 - D(G(z)))] \quad (4)$$

where λ is a hyperparameter that controls the relative importance of the L1 term.

IV. Results and Discussion

The performance of the proposed two-stage framework was rigorously evaluated using both quantitative and qualitative metrics to assess the effectiveness of speckle reduction and colourisation

A. *Evaluation Metrics*

A variety of metrics were used to assess the performance of both the denoising and colourisation stages, capturing aspects of both pixel-level accuracy and perceptual quality.

For Denoising (Reconstruction Quality)

- Mean Squared Error (MSE): A fundamental metric that calculates the average squared difference between the pixel values of the ground-truth image (I) and the reconstructed image (K). A lower MSE indicates a better reconstruction. It is defined as:

$$L_{MSE} = \frac{1}{N}\sum_{i=1}^{N} (y_i - \hat{y}_i)^2 \quad (5)$$

where m and n are the dimensions of the image.

- Peak Signal-to-Noise Ratio (PSNR): A widely used metric in image processing, PSNR measures the ratio of the maximum possible power of a signal to the power of the corrupting noise. It is expressed on a logarithmic decibel (dB) scale, and a higher value signifies better quality. PSNR is inversely related to MSE and is calculated as:

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

where MAX_I is the maximum possible pixel value of the image (e.g., 255 for an 8-bit grayscale image).

- Structural Similarity Index Measure (SSIM): Unlike PSNR and MSE, which are based on absolute errors, SSIM is a perceptual metric that evaluates image quality degradation based on changes in structural information. It compares local patterns of pixel intensities by incorporating luminance (l), contrast (c), and structure (s). The SSIM index ranges from -1 to 1, where 1 indicates perfect structural similarity. The formula for two image windows x and y is:

$$SSIM(x,y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)} \quad (7)$$

where μ is the mean, σ2 is the variance, σ_xy is the covariance, and c_1,c_2 are stabilisation constants.

For Colourisation (Generative Quality)

- Fréchet Inception Distance (FID): A standard metric for evaluating the quality of generative models. FID measures the distance between the feature distributions of a set of real images and a set of generated images, as extracted by a pre-trained InceptionV3 network. It captures both the fidelity and diversity of the generated samples. A lower FID score indicates that the two distributions are more similar, signifying higher-quality results.

- Learned Perceptual Image Patch Similarity (LPIPS): Also known as perceptual distance, LPIPS measures the similarity between two images by computing the distance between their feature representations extracted from deep layers of a pre-trained neural network. This metric has been shown to align remarkably well with human perceptual judgments of image similarity. A lower LPIPS score indicates that two images are more perceptually similar.

B. *Quantitative and Qualitative Results*

Quantitative Analysis

The performance of the first stage, the Denoising CAE, is presented in Table I. The model achieves a low training loss and high PSNR and SSIM values, indicating its strong capability to remove speckle noise while faithfully preserving the underlying image structure.

I. TABLE 1: PERFORMANCE OF DENOISING CAE

Metric	*Value Range*	
Training Loss (MSE)	0.007	0.015
PSNR (dB)	25	32
SSIM	0.75	0.88

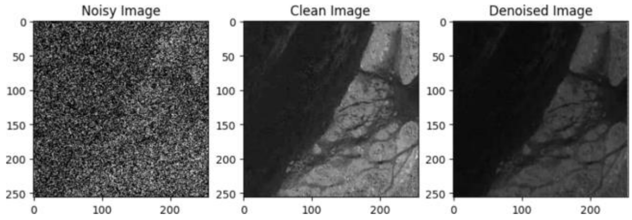
The performance of the pix2pix Colourisation GAN is presented in Table II. The table provides image quality metrics for the final output and the GAN's training losses, which provide insight into the adversarial training dynamics.

II. TABLE 2: PERFORMANCE OF PIX2PIX COLORIZATION GAN

Image Quality Metrics	
Metric	*Value*
SSIM	0.42
PSNR (dB)	18.5
FID	72.3
LPIPS	0.28
GAN Training Loss Components	
Loss Component	*Value*
Generator Total Loss	7.8
GAN Adversarial Loss	0.95
L1 Loss	0.068
Discriminator Loss	0.88

Qualitative Analysis

To complement the quantitative metrics, a qualitative analysis was performed by visually inspecting the outputs of the framework. A series of figures would present side-by-side comparisons for various scenes, showcasing the original noisy SAR input, the intermediate denoised output from the CAE, the final colourised output from the pix2pix GAN, and the corresponding ground-truth optical image.



These visual examples serve to illustrate the effectiveness of each stage. The output of the CAE clearly shows a significant reduction in the granular speckle pattern while retaining sharp edges and structural details. The final colourised output demonstrates the GAN's ability to generate realistic colours and textures for different land cover types, such as deep blues for water bodies, varied greens for vegetation, and greys and browns for urban structures and bare earth.
These visual results provide compelling evidence of the framework's ability to transform challenging, uninterpretable SAR data into visually intuitive and analytically valuable imagery.

C. *Discussion*

The quantitative results presented in Table I provide strong evidence for the efficacy of the Convolutional Autoencoder in the speckle reduction task. The achieved PSNR values, ranging from 25 dB to 32 dB, are indicative of a high-quality reconstruction with significantly reduced noise levels.
Similarly, the SSIM scores, which reach up to 0.88, confirm that the structural integrity of the images is well-preserved during the denoising process. The consistently low MSE training loss further suggests that the network converged successfully to a solution that accurately maps noisy inputs to their clean counterparts. The success of this first stage is a foundational enabler for the entire pipeline. By effectively removing the stochastic and corrupting influence of speckle, the CAE provides a normalised and stable input distribution to the subsequent colourisation stage. This dramatically simplifies the learning problem for the pix2pix GAN, which can then dedicate its full capacity to the already challenging task of cross-domain translation.

The performance of the colourisation stage, detailed in Table II, requires a nuanced interpretation. The pixel-based metrics, PSNR (18.5 dB) and SSIM (0.42), are relatively low. This is an expected outcome, as the goal of the GAN is not to reproduce the ground-truth optical image pixel-for-pixel, but to generate a colourisation consistent with the input SAR structure. Since colourisation is an ill-posed, one-to-many problem, the generated output may differ in fine details from the single ground-truth instance, leading to poor scores on these strict fidelity metrics. The perceptual metrics, FID (72.3) and LPIPS (0.28), offer a more meaningful assessment. These scores indicate that the distribution of generated images is perceptually and statistically close to the distribution of real optical images. The model is successfully learning the complex, high-level features, textures, and colour palettes that characterise natural scenes. The breakdown of the GAN losses further suggests that the adversarial training process reached a healthy equilibrium, avoiding common pitfalls like mode collapse.

Visual inspection of the output images corroborates the quantitative findings. The framework consistently produces colourised images with sharp edges, realistic textures, and semantically appropriate colours. However, a critical evaluation also reveals certain limitations. The colourisation task is inherently ambiguous, and the model can sometimes produce colours that are plausible but not factually correct for a specific location or time. For instance, it might colour a field green when it was fallow (brown) at the time of the optical image acquisition. This semantic leap is a fundamental challenge in any cross-domain translation task where one domain (SAR) contains strictly less information than the target domain (optical).

The proposed two-stage framework acts as a powerful domain normaliser, making raw SAR data from a challenging domain (noisy, single-channel) into one that is more aligned with human understanding and standard computer vision tools. By making SAR data more intuitive and optical-like, this framework dramatically lowers the barrier to entry for SAR data analysis. It allows human experts not trained in SAR interpretation to rapidly analyse scenes and unlocks the potential to apply a vast ecosystem of pre-trained computer vision models, for tasks like object detection and land-cover classification, directly to the enhanced SAR output, improving performance in a wide range of downstream applications.

V. FUTURE SCOPE

The main novelty of this work lies in the design and validation of a sequential, two-stage deep learning framework that explicitly decouples the tasks of denoising and colourisation [17]. Unlike singular approaches that attempt to solve both problems simultaneously, this modular pipeline employs specialised architectures for each stage: a Convolutional Autoencoder for speckle reduction and a conditional GAN for colourisation [9], [1]. This separation simplifies the learning objective for each network, allowing the CAE to focus on high-fidelity reconstruction and the GAN to dedicate its full capacity to the complex semantic inference of colour, resulting in a more robust and effective overall enhancement. The contributions of this research are summarised by a novel two-stage pipeline, where the design and implementation of an end-to-end framework systematically enhances SAR imagery by first performing speckle reduction and then colourisation, with rigorous validation using a hybrid suite of evaluation metrics. The denoising stage was assessed with reconstruction metrics (PSNR, SSIM), demonstrating high structural fidelity [14], while the colourisation stage was evaluated with perceptual metrics (FID, LPIPS), confirming the generation of realistic and visually plausible images that align well with human perception [3]. The framework successfully transforms noisy, monochromatic SAR data into clean, colourised images that are more intuitive for human analysts and compatible with standard computer vision algorithms, thereby increasing the accessibility and analytical value of SAR technology [1].

While the proposed framework demonstrates considerable success, there exists significant room for future research and improvement. The field of generative modelling is advancing rapidly [18]. Future work could explore replacing the pix2pix GAN with more recent and powerful architectures, such as denoising diffusion probabilistic models (DDPMs), for the colourisation stage [16]. Diffusion models have shown strong performance in image synthesis and may lead to even higher-fidelity and more diverse colourisations. Additionally, using attention mechanisms within the network architectures could help the models better capture long-range spatial dependencies, potentially improving the coherence of large-scale structures [4]. The current framework relies on a combination of adversarial and L1 loss. Future iterations could investigate the integration of a perceptual loss, such as LPIPS itself, directly into the generator's objective function [3], which would train the generator to minimise perceptual distance and may further improve visual quality while reducing reliance on the pixel-wise L1 term. A crucial next step is to quantitatively evaluate the impact of this enhancement pipeline on downstream analytical tasks. By using the colourised output as input for models trained on tasks like land-cover classification, object detection, or change detection, it would be possible to measure the concrete improvement in task-specific accuracy and demonstrate the tangible benefits of the proposed preprocessing [8]. This work focused on single-channel intensity SAR data. Another direction for future research is to adapt and extend the framework to handle more complex SAR modalities, such as polarimetric SAR (PolSAR) [19]. PolSAR data contains significantly richer information about the scattering properties and physical structure of objects on the ground. Using this additional information could enable more accurate and detailed colourisations, further closing the gap between SAR-derived products and true optical imagery [4].

VI. CONCLUSION

This paper has addressed the critical challenges of speckle noise and poor interpretability in Synthetic Aperture Radar imagery. A novel, two-stage deep learning framework was proposed, designed to systematically enhance SAR data for improved analysis. The first stage, a Convolutional Autoencoder, was shown to be highly effective at speckle reduction, achieving strong quantitative performance in terms of PSNR and SSIM, thereby preserving essential structural information. The second stage, a conditional GAN based on the pix2pix architecture, successfully translated the denoised grayscale images into realistic, full-colour representations. The quality of this translation was validated using modern perceptual metrics, FID and LPIPS, which demonstrated that the generated images are statistically and perceptually similar to real optical images. By decoupling the tasks of denoising and colourisation, the proposed framework offers a modular, robust, and effective solution that significantly enhances the visual quality and intuitive value of SAR imagery.

REFERENCES

1. G. Ji, Z. Wang, L. Zhou, Y. Xia, S. Zhong and S. Gong, "SAR Image Colorization Using Multidomain Cycle-Consistency Generative Adversarial Network," in IEEE Geoscience and Remote Sensing Letters, vol. 18, no. 2, pp. 296-300, Feb. 2021, doi: 10.1109/LGRS.2020.2969891.
2. I. Meraoumia, E. Dalsasso, L. Denis, R. Abergel, and F. Tupin, "Multitemporal speckle reduction with self-supervised deep neural networks," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5201914.
3. S. M. D'Souza, P. Balasubramanian, P. R. M, P. N. Reddy, P. K. A, and A. Shankar, "Interactive SAR Image Colorization Using Conditional GANs, ResNet and Generative Adversarial Refinement with Real-Time Insights," in *Proc. 6th Int. Conf. Mobile Computing and Sustainable Informatics (ICMCSI)*, Goathgaun, Nepal, 2025.
4. Z. Guo, H. Guo, X. Liu, W. Zhou, Y. Wang, and Y. Fan, "Sar2color: Learning Imaging Characteristics of SAR Images for SAR-to-Optical Transformation," *Remote Sens.*, vol. 14, 2022, Art. no. 3740, doi: 10.3390/rs14153740.
5. H. Imad, Z. Sara, M. Hajji, T. Yassine, and N. Abdelkrim, "Recent Advances in SAR Image Analysis Using Deep Learning Approaches: Examples of Speckle Denoising and Change Detection," in *Proc. 4th Int. Conf. Innovative Research in Applied Science, Engineering and Technology (IRASET)*, FEZ, Morocco, 2024, pp. 1–6, doi: 10.1109/IRASET60544.2024.10549456.
6. Z. Ji, L. Li, and H. Bi, "Deep Learning-Based Approximated Observation Sparse SAR Imaging via Complex-Valued Convolutional Neural Network," *Remote Sens.*, vol. 16, 2024, Art. no. 3850, doi: 10.3390/rs16203850.
7. B. Yonel, E. Mason, and B. Yazıcı, "Deep learning for passive synthetic aperture radar," *IEEE J. Sel. Topics Signal Process.*, vol. 12, pp. 90–103, 2018.
8. S. Zhang, L. Cui, Y. Zhang, T. Xia, Z. Dong, and W. An, "Research on Input Schemes for Polarimetric SAR Classification Using Deep Learning," *Remote Sens.*, vol. 16, no. 11, 2024, Art. no. 1826, doi: 10.3390/rs16111826.
9. C. Lin, C. Qiu, H. Jiang, and L. Zou, "A Deep Neural Network Based on Prior-Driven and Structural Preserving for SAR Image Despeckling," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 16, pp. 6372–6392, 2023, doi: 10.1109/JSTARS.2023.3292325.
10. B. Li, Y. Lu, W. Pang, et al., "Image Colorization using CycleGAN with semantic and spatial rationality," *Multimed. Tools Appl.*, vol. 82, pp. 21641–21655, 2023, doi: 10.1007/s11042-023-14675-9.
11. C. Wang, A. Mouche, P. Tandeo, J. Stopa, N. Longépé, G. Erhard, R. Foster, D. Vandemark, and B. Chapron, "Labeled SAR imagery dataset of ten geophysical phenomena from Sentinel-1 wave mode (TenGeoP-SARwv)," SEANOE, 2018, doi: 10.17882/56796.
12. P. Wang and V. M. Patel, "Generating high quality visible images from SAR images using CNNs," in *2018 IEEE Radar Conf. (RadarConf18)*, Oklahoma City, OK, USA, 2018, pp. 570–575, doi: 10.1109/RADAR.2018.8378622.
13. S. Dabhi, K. Soni, U. Patel, P. Sharma, and M. Parmar, "Virtual SAR: A Synthetic Dataset for Deep Learning based Speckle Noise Reduction Algorithms," IEEE Dataport, 2020, doi: 10.21227/asth-ra98.
14. J. Wang and K. Xu, "A Convolutional Neural Network SAR Image Denoising Algorithm Based on Self-Learning Strategies," *Appl. Sci.*, vol. 15, 2025, Art. no. 4786, doi: 10.3390/app15094786.
15. P. Wang, et al., "Generative adversarial network-based restoration of speckled SAR images," in *2017 IEEE 7th Int. Workshop Comput. Advances Multi-Sensor Adaptive Process. (CAMSAP)*, 2017, pp. 1–5.
16. M. V. Perera, et al., "SAR despeckling using a denoising diffusion probabilistic model," *IEEE Geosci. Remote Sens. Lett.*, vol. 20, 2023, pp. 1–5.
17. Q. Zhang and R. Sun, "SAR image despeckling based on convolutional denoising autoencoder," arXiv preprint arXiv:2011.14627, 2020.
18. S. Anwar, et al., "Image colorization: A survey and dataset," arXiv preprint arXiv:2008.10774, 2020.
19. A. Mestre-Quereda and J. M. Lopez-Sanchez, "Deep learning based speckle filtering for polarimetric SAR images. Application to Sentinel-1," *IEEE Trans. Geosci. Remote Sens.*, 2025.
20. R. Nobre, A. Rodrigues, R. Rosa, F. Medeiros, R. Feitosa, A. Estevão, and A. Barros, "GRSS SAR/PolSAR DATABASE," IEEE Dataport, 2017, doi: 10.21227/H28W2J.