

Deep Learning-Based SAR Super-Resolution for Accurate Water Body Mapping

Jung-Hoon Lee

AI Convergence Network

Ajou University

Suwon, Korea

geniuslee20@ajou.ac.kr

Tae-Yoon Kim

AI Convergence Network

Ajou University

Suwon, Korea

xodbsxogjs@ajou.ac.kr

Jongtae Lee

AI Convergence Network

Ajou University

Suwon, Korea

jtleee830@ajou.ac.kr

Min-Sik Kim

AI Convergence Network

Ajou University

Suwon, Korea

qq01025@ajou.ac.kr

Yu-Jeong Ahn

AI Convergence Network

Ajou University

Suwon, Korea

dbwjd5825@ajou.ac.kr

Habeen Oh

AI Convergence Network

Ajou University

Suwon, Korea

habeen0727@ajou.ac.kr

Jae-Hyun Kim

Electrical and Computer Engineering

Ajou University

Suwon, Korea

jkim@ajou.ac.kr

Abstract—Fine-resolution Synthetic Aperture Radar (SAR) data is crucial for many geospatial applications but is often limited by high cost and accessibility. This study introduces a novel deep learning framework to generate high-resolution SAR imagery, stylistically equivalent to COSMO-SkyMed, from freely available Sentinel-1 data. To mitigate inherent sensor discrepancies, our comprehensive pipeline performs precise co-registration, radiometric calibration, and domain adaptation to bridge the spectral and structural differences between the C-band and X-band inputs. The core of our framework is a multi-scale conditional GAN, inspired by Pix2pixHD, which uses a coarse-to-fine generator and multi-scale discriminators with a composite loss function to reconstruct detailed SAR textures. Our model not only surpasses existing baselines on key image quality metrics (SSIM, PSNR, DISTS) but also demonstrates tangible practical value. When integrated into a downstream water body detection pipeline, the super-resolved images achieved an Intersection-over-Union (IoU) score exceeding 70%, enabling significantly more precise mapping than possible with original low-resolution data. This work validates deep learning-based super-resolution as a scalable and cost-effective pathway to produce high-fidelity SAR data, effectively democratizing access for critical applications like flood monitoring and disaster assessment.

Index Terms—SAR super-resolution, COSMO-SkyMed, Sentinel-1, water body detection, deep learning, remote sensing.

I. INTRODUCTION

Satellite-based remote sensing is a critical tool for observing the Earth's surface and atmosphere, supporting diverse applications ranging from environmental monitoring to urban planning. Among the various sensor types, Synthetic Aperture Radar (SAR) stands out due to its all-weather, day-and-night imaging capabilities, allowing for the detection of subtle surface changes and the generation of high-contrast imagery even under low-visibility conditions [1]. These advantages make SAR essential for tasks such as monitoring ice sheets, detecting pollution, and assessing infrastructure [2]. Despite its benefits, integrating SAR data across different satellite platforms presents significant challenges. Variations in frequency bands,

spatial resolutions, and acquisition parameters introduce inconsistencies that hinder seamless analysis. While Sentinel-1 provides extensive coverage, many applications—such as accurate water body detection for flood management and hydrological modeling—require the finer spatial detail available from higher-resolution systems like COSMO-SkyMed. Bridging this resolution and domain gap in a cost-effective manner remains a pressing research goal. Recent advances in deep learning, especially in Convolutional Neural Networks (CNNs) [3], Generative Adversarial Networks (GANs) [4] [5], and Transformer-based architectures [6] [7], have significantly advanced the field of image super-resolution. However, most work to date has focused on optical imagery, whereas SAR data presents unique challenges. Destructive speckle noise can be misinterpreted as fine texture by generative models, and significant domain shifts between frequency bands can hinder model generalization. Furthermore, the scale and diversity of SAR datasets impose computational burdens, necessitating robust, data-driven methods that incorporate band normalization, speckle noise reduction, and domain adaptation. In response, this paper proposes a novel deep learning framework to generate high-resolution COSMO-SkyMed-like SAR images from Sentinel-1 inputs. We leverage GANs and Transformer-based models for their complementary strengths in texture synthesis and modeling long-range dependencies. Our approach explicitly addresses key SAR-specific challenges, including speckle suppression, cross-band discrepancies, and large-scale data tiling. The main contributions of this study are as follows: (1) A systematic and reproducible end-to-end data preparation pipeline to harmonize Sentinel-1 and COSMO-SkyMed data, addressing critical inconsistencies; (2) A novel hybrid deep learning architecture that synergizes GANs and Transformers, specifically tailored to reconstruct SAR textures while preserving structural integrity; (3) Comprehensive validation of the proposed framework through both objective metrics and perceptual quality assessment, confirming the generation of

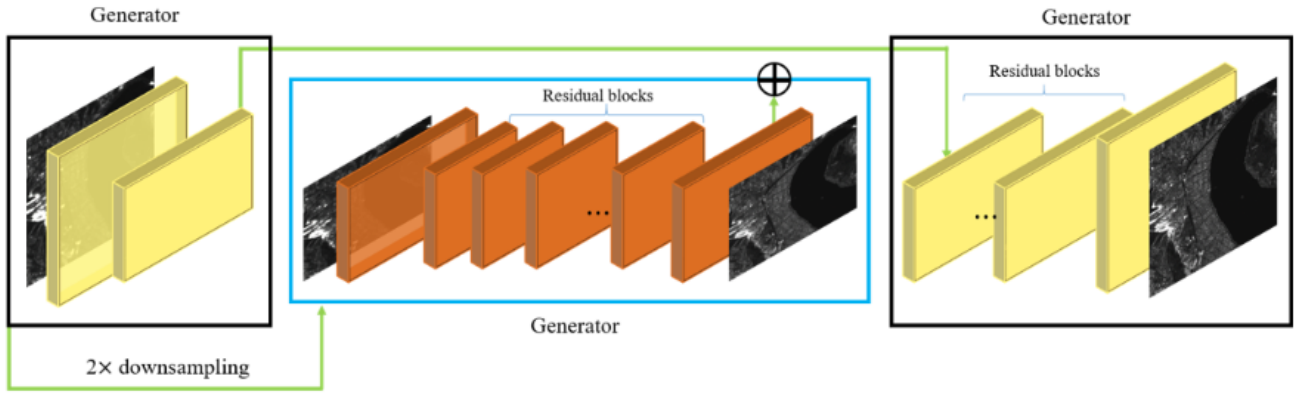


Fig. 1. Proposed Model's Architecture

high-fidelity SAR images; and (4) A practical demonstration of the framework's real-world utility, where the super-resolved outputs significantly improve the accuracy of surface water body detection for environmental monitoring.

II. PROPOSED METHODS

A. Pre-processing

Our study employs a systematic, multi-stage pre-processing pipeline to harmonize Sentinel-1 and COSMO-SkyMed data, ensuring the deep learning model focuses on meaningful texture synthesis rather than compensating for sensor-specific physical discrepancies. The process begins with precise geometric co-registration to a common geodetic reference frame, followed by radiometric calibration to convert pixel values into normalized radar cross-sections for consistent backscatter measurements. To further align the statistical properties, we then apply histogram matching as a final radiometric harmonization step. Subsequently, we address SAR-specific artifacts by applying a light speckle filter (e.g., a 3x3 Lee filter) and aligning all data to a consistent polarization channel (e.g., VV). This critical early-stage filtering is intended to reduce inherent speckle noise before it can be amplified by the generative network, which could otherwise mistake noise for fine-grained texture. To bridge the significant domain gap between Sentinel-1, C-band and COSMO-SkyMed, X-band, then employ a dedicated domain adaptation module that uses a feature-level transformation technique [e.g., a CycleGAN-based approach] to translate the echo characteristics of the input imagery to better resemble the target domain. Finally, the fully preprocessed, full-scene images are automatically tiled into smaller patches and augmented with random flipping and rotation to generate a large-scale, robust training dataset. The entire pipeline is efficiently implemented using MATLAB.

B. Deep Learning Model

At the core of our framework is a conditional Generative Adversarial Network (CGAN) specifically designed for SAR super-resolution, inspired by the Pix2pixHD architecture. The Generator network follows a multi-scale, coarse-to-fine strategy. A global generator branch first processes downsampled

inputs to learn large-scale context and bridge the domain gap between C-band and X-band data, while a local enhancer branch refines fine-grained details at the target resolution. This dual-path design, equipped with skip connections and instance normalization, ensures both contextual coherence and the preservation of subtle local textures. To further address cross-sensor discrepancies at the feature level, we integrate an early feature alignment layer, implemented as a set of residual blocks, at the beginning of the generator. In parallel, a multi-scale Discriminator network evaluates the generated imagery. Following the PatchGAN paradigm, it assesses the realism of overlapping local patches rather than the entire image. Critically, to make the discriminator SAR-aware, we augment its input with additional channels representing speckle statistics specifically, local mean and variance maps, enabling it to enforce consistency in SAR-specific texture. The entire network is trained end-to-end with a composite loss function, defined as a weighted sum of an adversarial loss, a pixel-wise L1 reconstruction loss, a VGG-based perceptual loss, and a Structural Similarity Index (SSIM) loss. This multi-component loss guides the generator to produce high-fidelity, radiometrically consistent outputs that are both visually plausible and quantitatively accurate.

C. Water Edge Detection

To validate the practical utility of our super-resolved imagery, we implemented a deterministic image processing pipeline for automated water body segmentation. This approach deliberately avoids a secondary deep learning model to transparently demonstrate the quality enhancement provided by our super-resolution framework itself, showing that even classic, interpretable methods can achieve high accuracy on the enhanced data. The pipeline begins with contrast enhancement via histogram equalization to improve feature visibility. It then applies Otsu's global thresholding with binary inversion to segment the image, operating on the common assumption that calm water bodies exhibit low backscatter and appear as dark regions. The resulting binary masks are refined using morphological opening and closing operations to remove noise artifacts and fill internal gaps, creating contiguous water

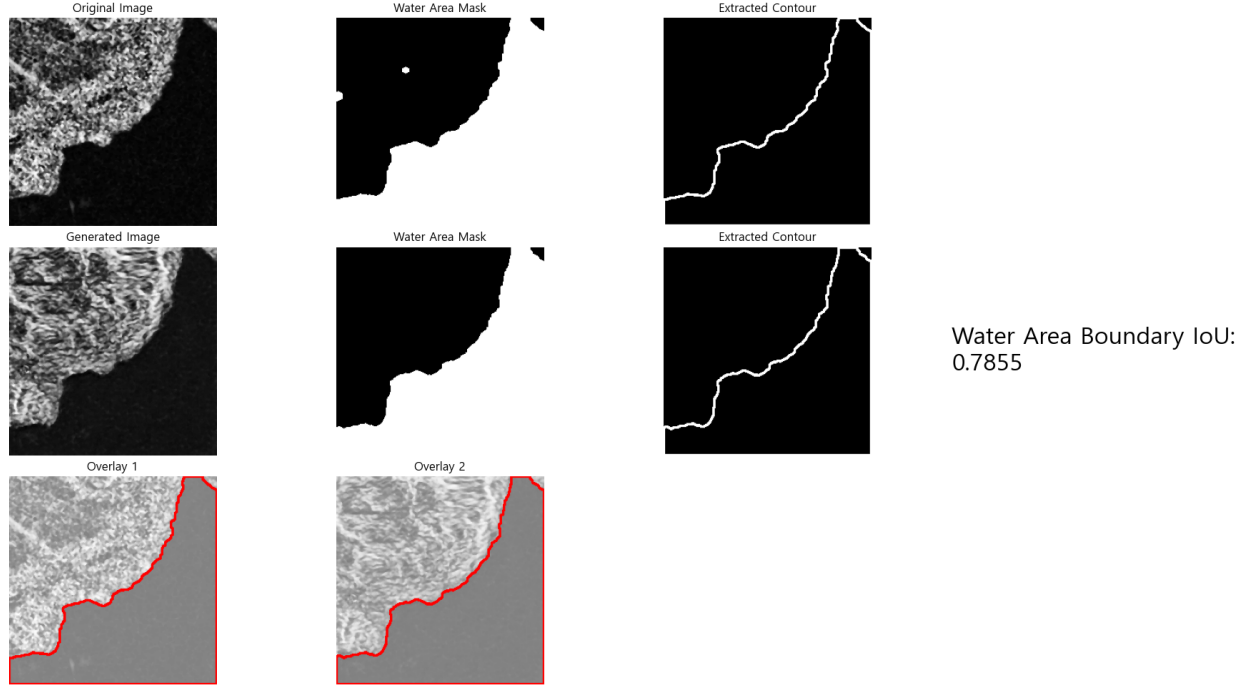


Fig. 2. Comparison of water body detection between original and generated images.

regions. From these clean masks, contours are extracted and filtered based on an area threshold to discard insignificant fragments, identifying the primary water body boundaries. To quantitatively assess performance, we compute the Intersection over Union (IoU) between the mask predicted from our super-resolved output and a reference mask generated from the ground-truth high-resolution imagery. For qualitative analysis, a multi-panel visualization is generated, displaying the input image, the final binary mask, and the extracted contour overlaid on the original image, alongside the calculated IoU score.

D. Indexes

In evaluating the proposed cross-band SR framework, we employ a triplet of quality indices—PSNR, SSIM, and, DISTS [8]. Structural Similarity Index (SSIM) offers a window-based assessment of luminance, contrast, and gradient consistency. This enables us to verify that large-scale geometric relationships—such as building footprints, transportation grids, and shoreline edges—remain intact even when speckle noise or minor misregistrations are present. Mathematically, it can be expressed as

$$SSIM(x, y) = [l(x, y)^\alpha] \cdot [c(x, y)^\beta] \cdot [s(x, y)^\gamma]. \quad (1)$$

where l , c , and s represent the luminance, contrast, and structure terms computed over local windows, x and y denote the image patches under comparison, and the exponents α , β , and γ weight the relative importance of each component.

Peak-Signal-to-Noise Ratio (PSNR), a logarithmic measure that compares the peak possible signal power with the power of distorting noise. Because PSNR is derived directly from the

mean-squared error, it provides a straightforward, if somewhat coarse, indication of reconstruction fidelity. It is computed as

$$PSNR = 10 \log_{10} \left(\frac{I_{\max}^2}{MSE} \right), \quad MSE = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2. \quad (2)$$

where $I_{\max}$ is the peak possible back-scatter intensity after normalization, typically 1, N is the number of pixels, y_i is the reference COSMO-SkyMed value, and $\hat{y}_i$ is the corresponding super-resolved prediction. A smaller MSE—and therefore a larger PSNR—indicates closer numerical agreement.

Deep Image Structure and Texture Similarity (DISTS) is adopted as the primary perceptual metric because it correlates with human judgements far better than traditional, hand-crafted indices. Built on the feature space of a pre-trained VGG-16 network, DISTS quantifies how closely a super-resolved (SR) patch $\hat{y}$ matches its reference y in both structural alignment and textural richness. Formally, the score is defined as

$$DISTS(y, \hat{y}) = 1 - \sum_{\ell} w_{\ell} [S_{\ell}(y, \hat{y}) + T_{\ell}(y, \hat{y})], \quad (8)$$

$$S_{\ell}(y, \hat{y}) = \frac{2 \langle \phi_{\ell}(y), \phi_{\ell}(\hat{y}) \rangle}{\|\phi_{\ell}(y)\|_2^2 + \|\phi_{\ell}(\hat{y})\|_2^2}, \quad (9)$$

$$T_{\ell}(y, \hat{y}) = \frac{2 \sigma_{\ell}(y) \sigma_{\ell}(\hat{y})}{\sigma_{\ell}^2(y) + \sigma_{\ell}^2(\hat{y})}. \quad (10)$$

S_{ℓ} represents the structure similarity at layer ℓ , while T_{ℓ} measures texture similarity by matching the channel-wise

variances of the paired feature maps. Here, $\phi_\ell(\cdot)$ denotes the VGG activation extracted at layer ℓ , and $\sigma_\ell(\cdot)$ is the corresponding standard deviation, capturing texture energy. And w_ℓ are non-negative weights obtained from large-scale human-preference studies.

III. EXPERIMENTAL RESULTS

TABLE I
COMPARING INDEXES

Index	SSIM	PSNR	DISTS
CGAN	0.5720	18.50	0.0941
CycleGAN	0.5942	20.14	0.0756
Pix2pixHD	0.6245	21.62	0.0642

Our proposed super-resolution framework was comprehensively evaluated through both quantitative metrics and a practical application, with the results summarized in Table 1 and Figure 2. In direct comparison with widely used GAN baselines, our model demonstrates superior performance in image reconstruction. As shown in Table 1, it achieves the highest scores in traditional metrics like SSIM and PSNR, and more importantly, the best score in the perceptually-aligned DISTS metric. This indicates that our generated images are not only quantitatively accurate but also visually more realistic in texture and structure than those from competing methods. This enhanced image fidelity directly translates to improved performance in the downstream task of water body segmentation. Qualitatively, the super-resolved images in Figure 2 exhibit markedly sharper and more well-defined water boundaries compared to the original Sentinel-1 inputs. This visual improvement is validated quantitatively, with the extracted water masks achieving high IoU scores ranging from 0.73 to 0.78. Ultimately, these combined results empirically validate that our framework effectively enhances low-resolution SAR data, providing tangible and high-value outputs for subsequent geospatial analysis tasks.

IV. CONCLUSION

This study introduced a deep learning-based super-resolution framework that effectively transforms widely available, low-resolution SAR imagery into high-resolution outputs, addressing the high acquisition costs and technical challenges associated with fine-resolution data. Our model, built upon a SAR-enhanced Pix2pixHD architecture, not only demonstrates superior quantitative performance over baseline methods across SSIM, PSNR, and DISTS metrics but also generates visibly superior outputs with sharper details and more realistic textures. The practical utility of this high-fidelity reconstruction was validated through water body detection experiments, where our method achieved Intersection over Union (IoU) scores exceeding 70%, confirming its ability to preserve structurally critical features for geospatial analysis.

Future work will focus on two key directions: (1) enhancing fine-detail reconstruction and robustness by exploring more advanced generative architectures and loss functions, and (2) improving model generalizability by curating a larger, more

heterogeneous training dataset covering diverse geographic conditions. Furthermore, we plan to adapt and validate our framework for more demanding applications, such as post-disaster infrastructure monitoring and change detection. In conclusion, this work establishes deep learning-based super-resolution as a robust and cost-effective force multiplier, transforming low-resolution SAR data into actionable, high-fidelity geospatial intelligence and paving the way for a new era of accessible Earth observation.

ACKNOWLEDGMENT

This research was supported by the Institute of Information & Communications Technology Planning & Evaluation (IITP) grant funded by the Korea Government (MSIT) (No.RS-2024-00396992 Development of Cube Satellites Based on Core Technologies in Low Earth Orbit Satellite Communications)

REFERENCES

- [1] A. Tsokas, M. Rysz, P. M. Pardalos, and K. Dipple, "Sar data applications in earth observation: An overview," *Expert Systems with Applications*, vol. 205, p. 117342, 2022.
- [2] X. Zhou, N.-B. Chang, and S. Li, "Applications of sar interferometry in earth and environmental science research," *Sensors*, vol. 9, no. 3, pp. 1876–1912, 2009.
- [3] K. O'shea and R. Nash, "An introduction to convolutional neural networks," *arXiv preprint arXiv:1511.08458*, 2015.
- [4] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," *Advances in neural information processing systems*, vol. 27, 2014.
- [5] X. Wang, K. Yu, S. Wu, J. Gu, Y. Liu, C. Dong, Y. Qiao, and C. Change Loy, "Esrgan: Enhanced super-resolution generative adversarial networks," in *Proceedings of the European conference on computer vision (ECCV) workshops*, 2018, pp. 0–0.
- [6] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention is all you need," *Advances in neural information processing systems*, vol. 30, 2017.
- [7] R. Xia, J. Chen, Z. Huang, H. Wan, B. Wu, L. Sun, B. Yao, H. Xiang, and M. Xing, "Crtranssar: A visual transformer based on contextual joint representation learning for sar ship detection," *Remote Sensing*, vol. 14, no. 6, p. 1488, 2022.
- [8] U. Sara, M. Akter, and M. S. Uddin, "Image quality assessment through fsim, ssim, mse and psnr—a comparative study," *Journal of Computer and Communications*, vol. 7, no. 3, pp. 8–18, 2019.