



# Sar-Optical Image Matching Enhancement Using Conditional Generative Adversarial Networks



Aigerim Sultanbekova<sup>1</sup>, Timo Balz<sup>1</sup>

<sup>1</sup>State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing (LIESMARS), Wuhan University, China; aigerims10.29@gmail.com, balz@whu.edu.cn

## Abstract

Synthetic Aperture Radar (SAR) and optical imagery offer complementary advantages in remote sensing, but their integration for accurate data fusion remains a challenging task due to inherent differences in their characteristics. This paper investigates the use of Generative Adversarial Networks (GANs) for improving SAR-optical image matching, focusing on the potential of conditional GANs (cGANs) for generating optical-like images from SAR data. We compare traditional feature-based methods, such as keypoint matching with the Scale Invariant Feature Transform (SIFT) and Random Sample Consensus (RANSAC) algorithms, shape-based contour matching, and pixel-wise absolute difference, to the results achieved by using cGAN-generated optical images. Our results indicate that matching SAR with optical images results in lower similarity scores, with a Structural Similarity Index (SSIM) score of 0.10 and a pixel-wise absolute difference score of 0.67. In contrast, when matching optical images with generated optical images from SAR data, the SSIM score improves to 0.20 and pixel wise absolute difference increases to 0.81. These findings suggest that cGANs significantly enhance the registration and matching of SAR and optical images by generating high-quality pseudo-optical images, thereby improving feature matching and pixel level alignment. This research demonstrates the potential of deep learning models in overcoming the challenges of multi-sensor data fusion, providing a robust framework for applications in land-use mapping, change detection, and environmental monitoring.

## Introduction

SAR and optical imagery offer complementary advantages in remote sensing applications. However, their integration remains challenging due to inherent differences in imaging mechanisms, radiometric properties, and geometric characteristics.

- Key Challenges:**
- Differences in imaging mechanisms and radiometric properties
  - Geometric and temporal variations between sensors
  - Sensitivity of SAR to noise and atmospheric conditions
  - Limited consistent features for reliable matching
  - Traditional methods struggle with cross-domain alignment

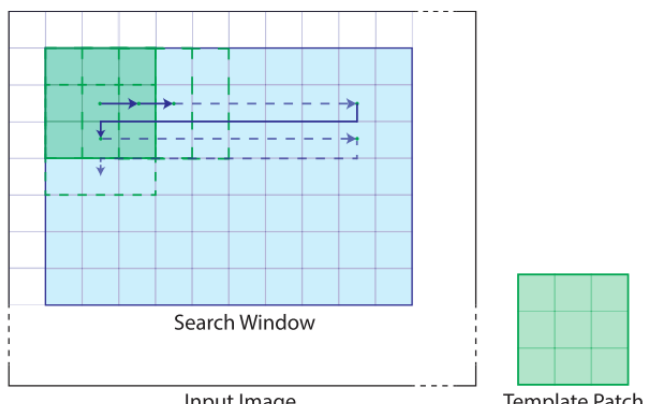


Figure 1: Area-Based (pixel intensity)

(Hughes 2020; Liao et al., 2020; Somervold et al., 2023)

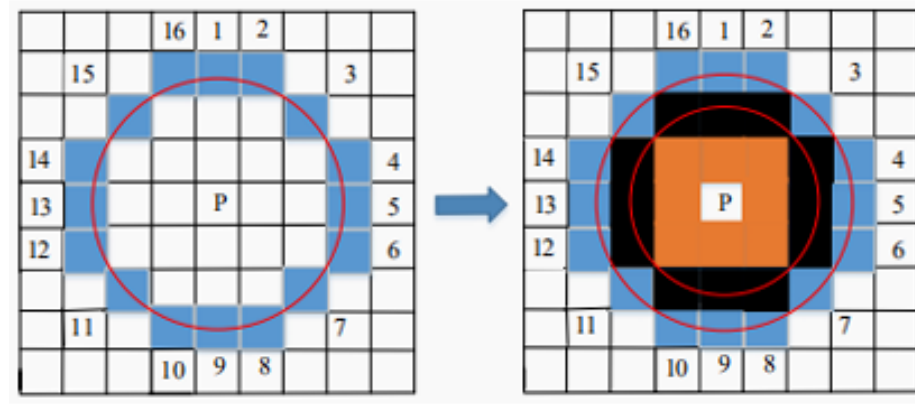


Figure 2: Feature-Based (descriptors as FAST, SIFT, ORB ect)

## Objectives

- To investigate the challenges of SAR-optical image fusion caused by the inherent differences in their sensor characteristics.
- To evaluate the effectiveness of conditional Generative Adversarial Networks (cGANs) in generating optical-like images from SAR data for improved image registration and matching.
- To compare the performance of traditional image matching techniques (e.g., SIFT with RANSAC, shape-based contour matching, and pixel-wise absolute difference) with cGAN-generated image matching results.
- To assess the improvement in structural similarity and pixel-wise alignment when using generated pseudo-optical images in SAR-optical fusion.
- To demonstrate the potential of deep learning-based approaches for enhancing multi-sensor data fusion applications in land-use mapping, change detection, and environmental monitoring.

## Materials and Methods

- Dataset & Evaluation Metrics**
- Dataset: QXS-SAROPT high-resolution SAR-optical image pairs
  - Evaluation Metrics:
    - Structural Similarity Index (SSIM)
    - Pixel-wise Absolute Difference
    - Feature Matching Scores with SIFT and RANSAC

This study compares a traditional image matching pipeline with a GAN-based deep learning approach to improve SAR-optical image registration. In the traditional approach, SAR and optical images were matched directly using feature-based methods such as SIFT and ORB descriptors. RANSAC was applied to eliminate outlier keypoints, and affine transformation was used for geometric correction. While effective for similar modalities, this method struggles when applied to SAR-optical pairs due to significant differences in imaging geometry and radiometric properties, resulting in limited matching accuracy.

At the heart of GANs is a two-player game between a Generator (G) and a Discriminator (D) (Fig.3). The Generator tries to produce realistic data, while the Discriminator aims to distinguish between real and fake data.

The GAN-enhanced approach addresses this challenge by translating SAR images into optical-like representations using a conditional Generative Adversarial Network (cGAN). This translation helps bridge the modality gap before performing feature matching (Fig.4).

The generator network follows a UNet architecture, leveraging an encoder-decoder structure with skip connections to preserve spatial and semantic details. These skip connections enable fine-grained feature aggregation, essential for high-resolution image generation.

The discriminator network acts as a binary classifier, distinguishing between real and generated optical images. It evaluates the realism of generated outputs and guides the generator through adversarial loss, encouraging the synthesis of visually convincing images.

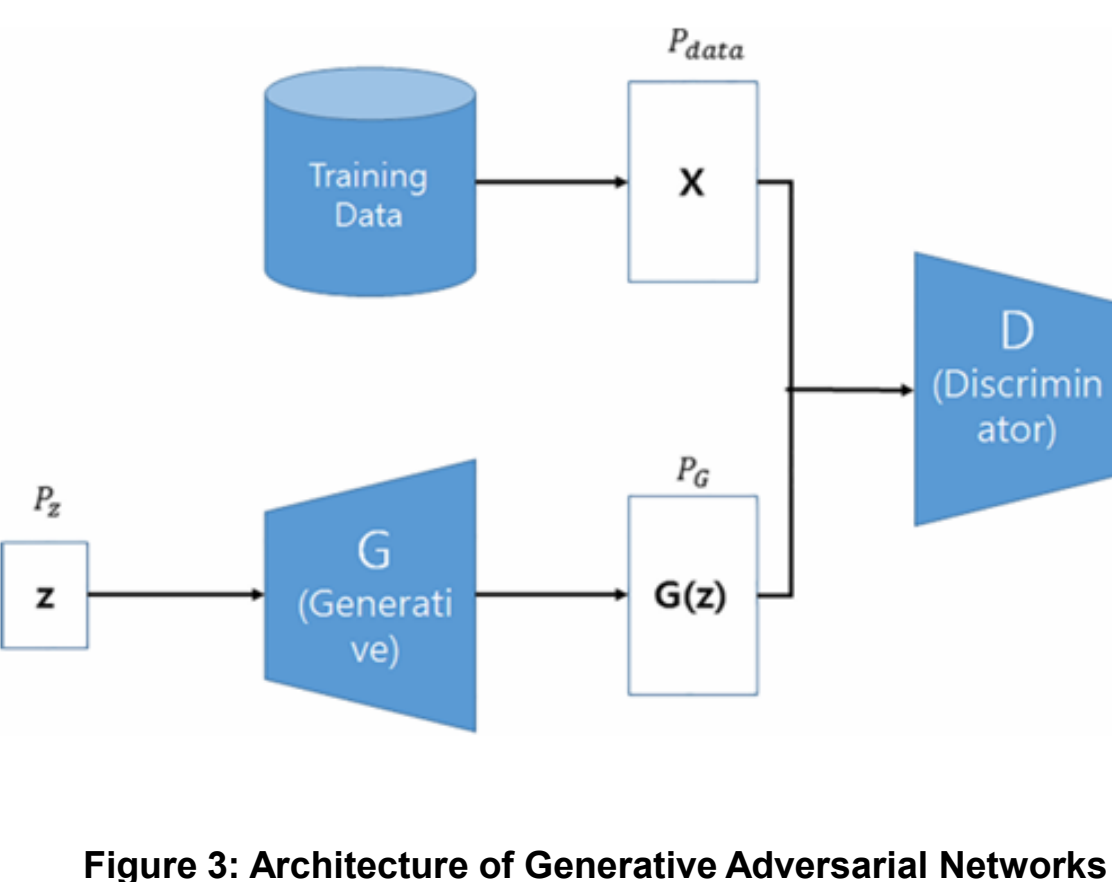


Figure 3: Architecture of Generative Adversarial Networks

This adversarial process is formalized as a min-max optimization problem through the Formula 1:

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

- $D(x|y)$  is the discriminator's output for real data  $x$ ,
- $G(z|y)$  is the generator's output conditioned on input  $y$ ,
- $p_{data}$  and  $p_z$  are distributions of real data and noise

During training, SAR images are fed as conditional inputs to guide the generation process. The model is optimized by minimizing adversarial loss while retaining structural information through skip connections. As a result, the generated pseudo-optical images exhibit enhanced visual similarity and spatial alignment with real optical data, leading to improved feature correspondence and more accurate registration.

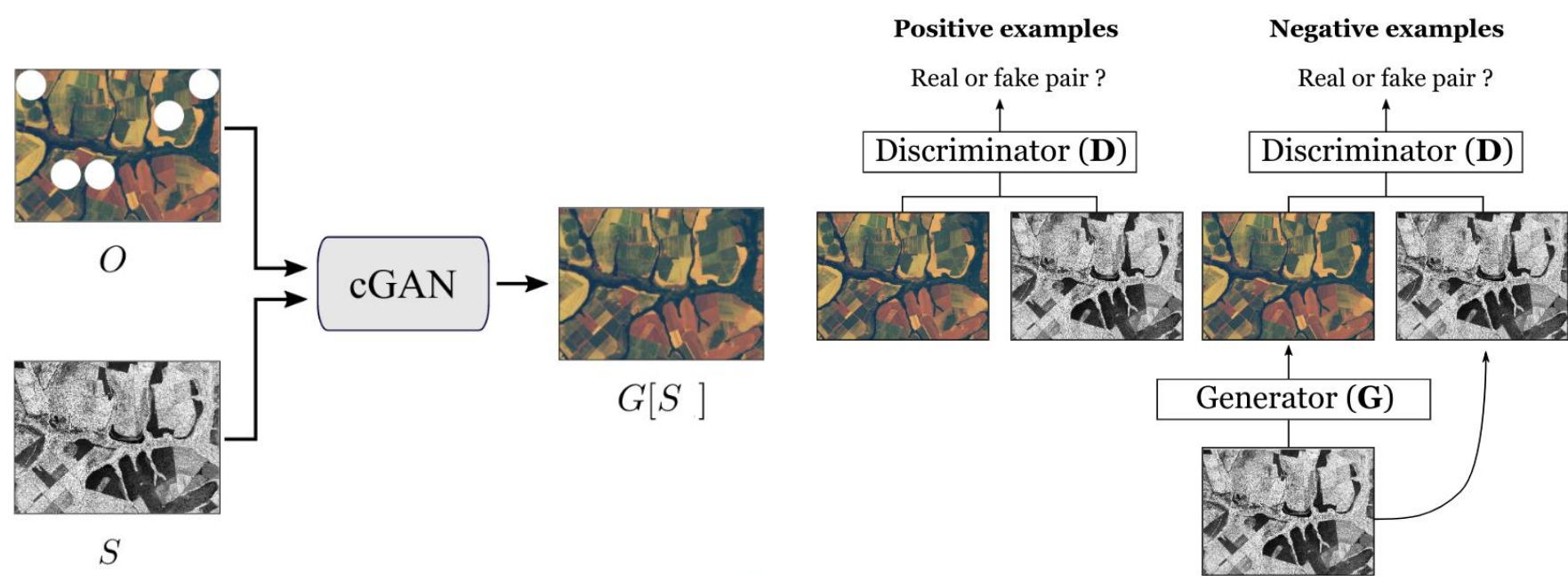


Figure 4: Architecture of SAR2OPT image translation conditional GAN

## Results

**Traditional SAR-Optical Matching Approach.** In the traditional pipeline, image registration was performed directly between SAR and optical images using feature-based methods such as SIFT and ORB descriptors. RANSAC was applied to eliminate mismatches, followed by affine transformation for geometric correction. Despite being widely used, this method struggled due to the inherent modality differences between SAR and optical data.

In the Table 1 matching evaluation was shown. SSIM Score: 0.10, indicating low structural similarity between SAR and optical images. Pixel-wise Absolute Difference: 0.67, reflecting poor alignment at the pixel level. Feature Matching Score: 0.94, suggesting the presence of many unmatched or incorrect keypoints. Visual Assessment: Misaligned structures and inconsistent feature correspondence due to modality gap.

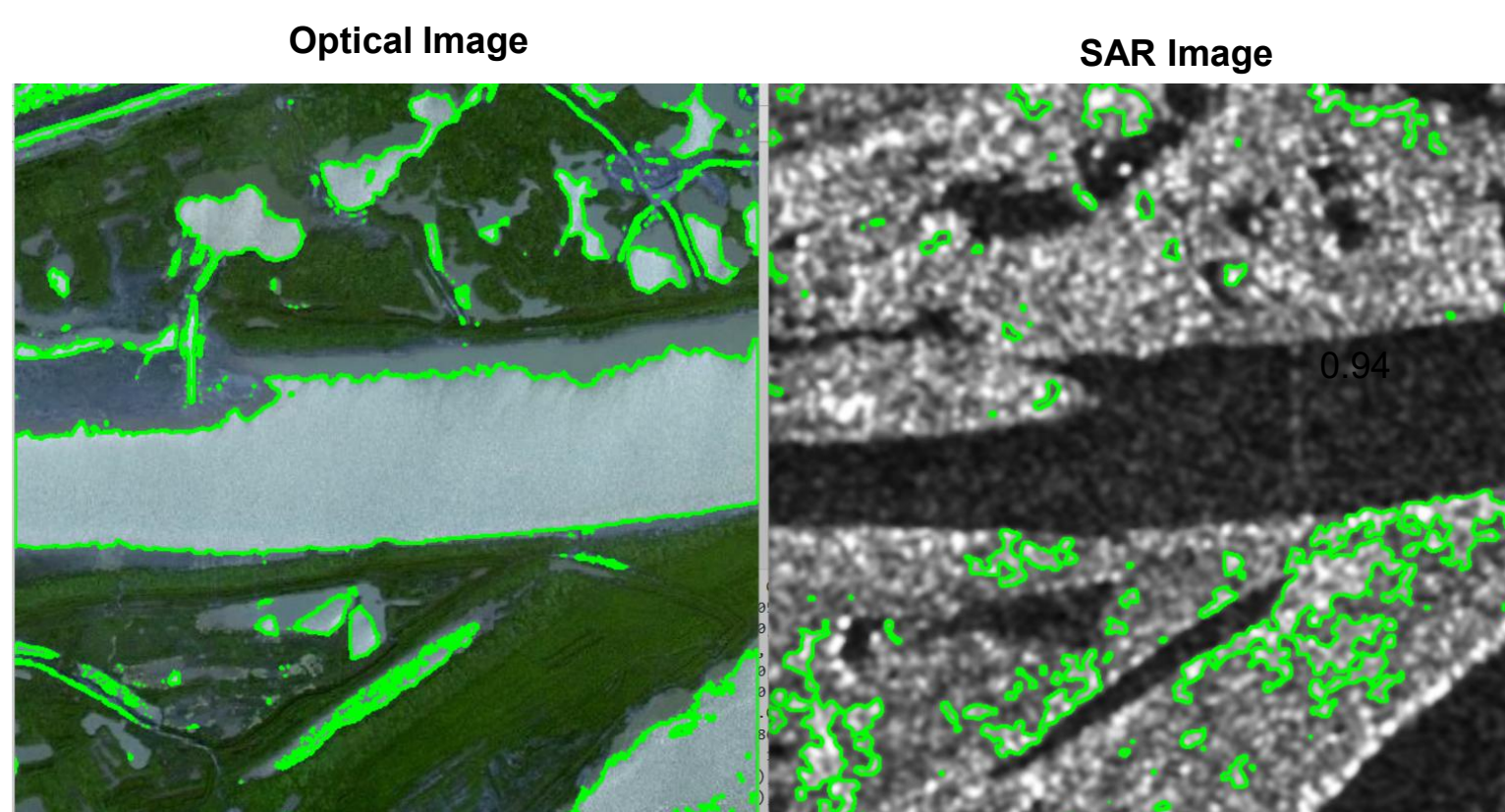


Figure 5: Shape and contour-based matching

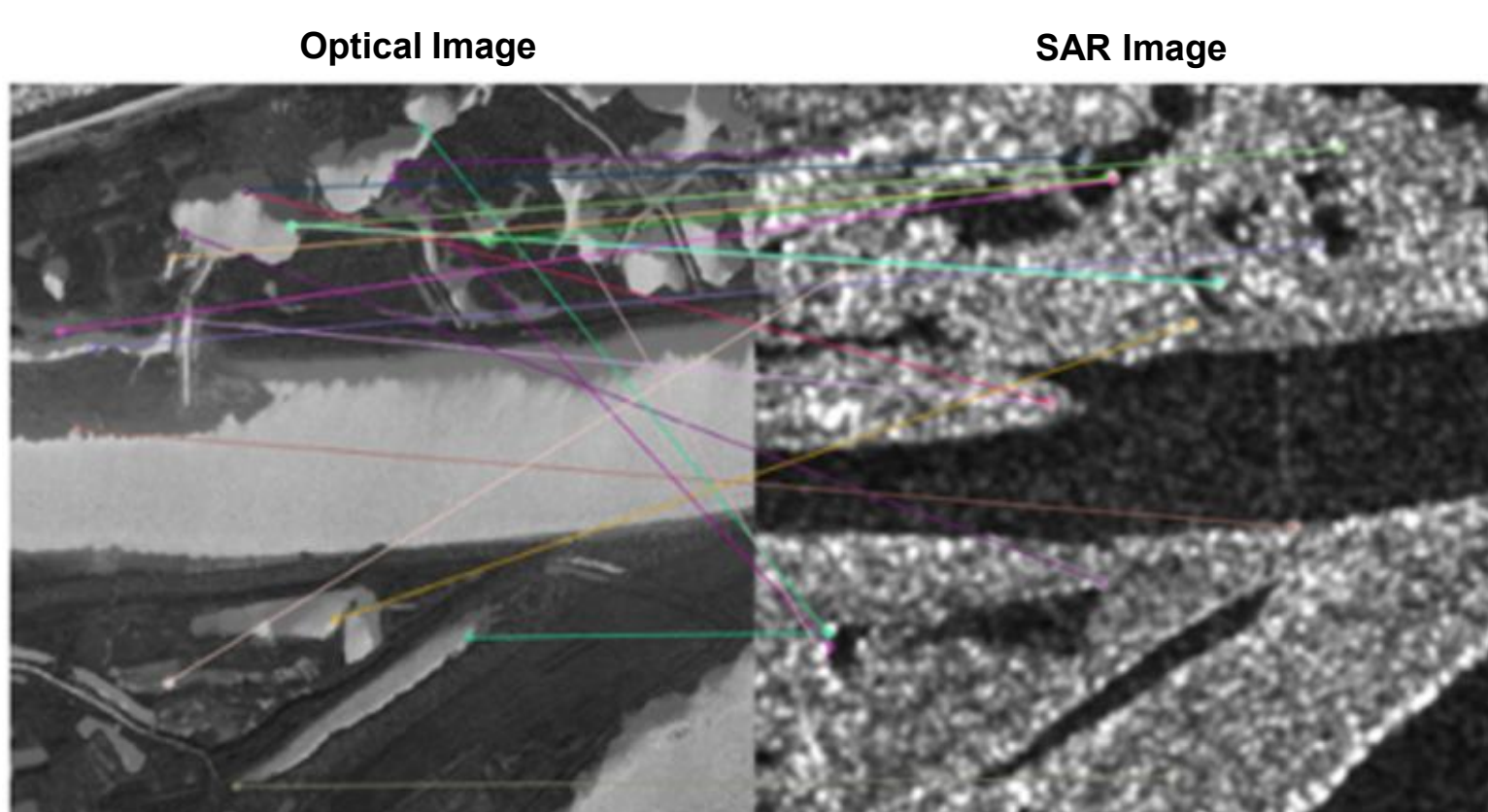


Figure 6: Feature based: SIFT and RANSAC

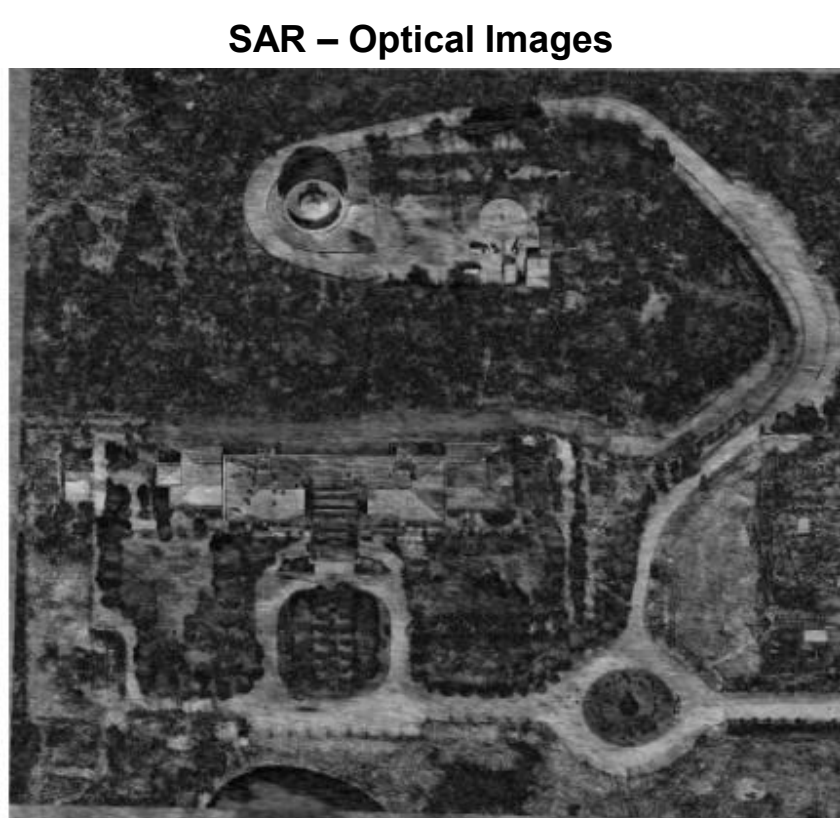


Figure 7: Pixel-wise Absolute Difference

| Method                           | Evaluation                                |
|----------------------------------|-------------------------------------------|
| Shape and contour-based matching | Matching score 0.94                       |
| Feature based: SIFT and RANSAC   | Structural similarity indices (SSIM) 0.10 |
| Pixel-wise Absolute Difference   | Score 0.67                                |

Table 1: Evaluation of Optical-SAR matching

In the Figure 8 the top row presents the outputs of a conditional GAN model applied to SAR-to-Optical image translation using two encoder-decoder architectures. From left to right: the original SAR input image, the generated optical image using a standard U-Net architecture, and the generated image using a U-Net with skip connections. The inclusion of skip connections allows for better preservation of spatial details and texture continuity. The bottom row shows zoomed-in regions of the corresponding images above, emphasizing differences in gradient sharpness and structural clarity. Visual inspection reveals that the U-Net with skip connections produces smoother gradients and more realistic features, indicating improved translation performance compared to the standard U-Net.

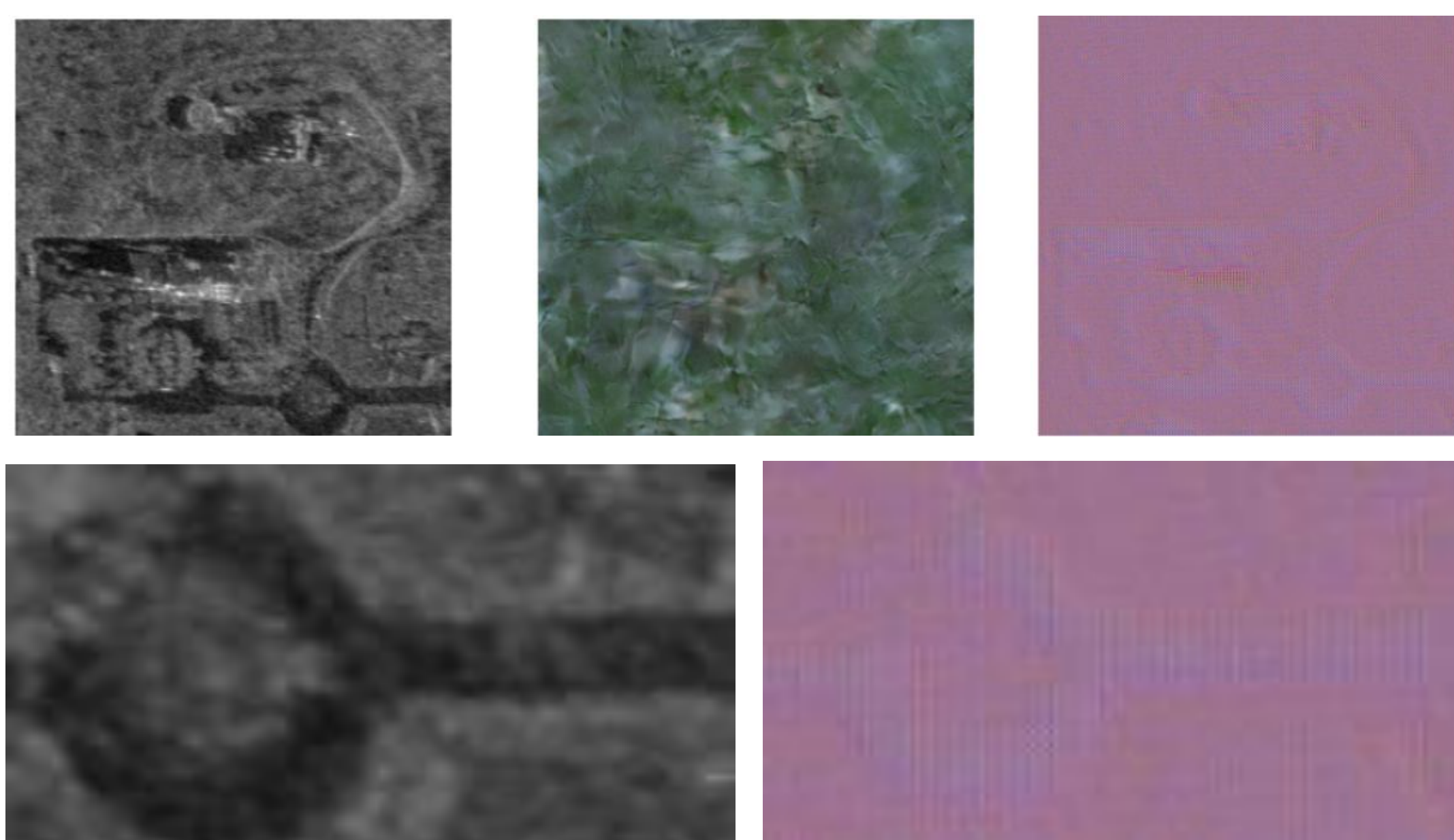


Figure 8: Top: Outputs of cGANs with two different Encoder-Decoder Architecture. Left to Right: SAR image, Standard UNet Generated Image, UNet with Skip Connections Generated Image. Bottom: Zoomed images to see the gradient. Left to Right: SAR image, UNet with Skip Connections Generated Image

The proposed GAN-based method significantly improved the quality of image matching by translating SAR images into pseudo-optical images using a conditional GAN with a UNet-based generator. Based on results of matching in Table 2 the generated images closely resembled the target optical images, enabling more accurate feature matching and alignment. SSIM Score: 0.20, representing an improvement in structural similarity. Pixel-wise Absolute Difference: 0.81, we can see improvement in image alignment quality. Feature Matching: Noticeably more consistent keypoint correspondence and higher confidence in matches. Visual Assessment: Enhanced spatial alignment and realistic image synthesis led to better fusion performance.

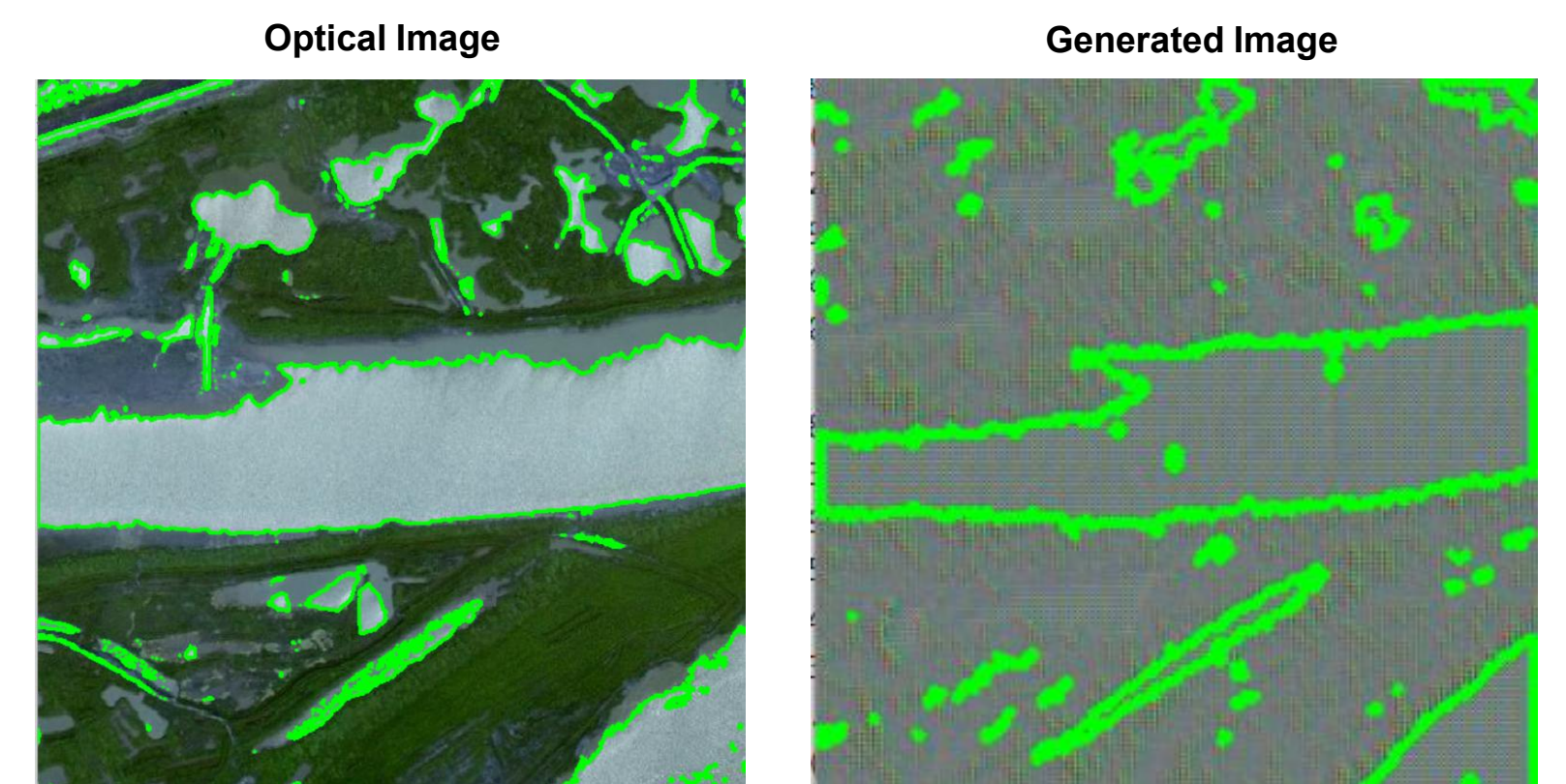


Figure 9: Shape and contour-based matching

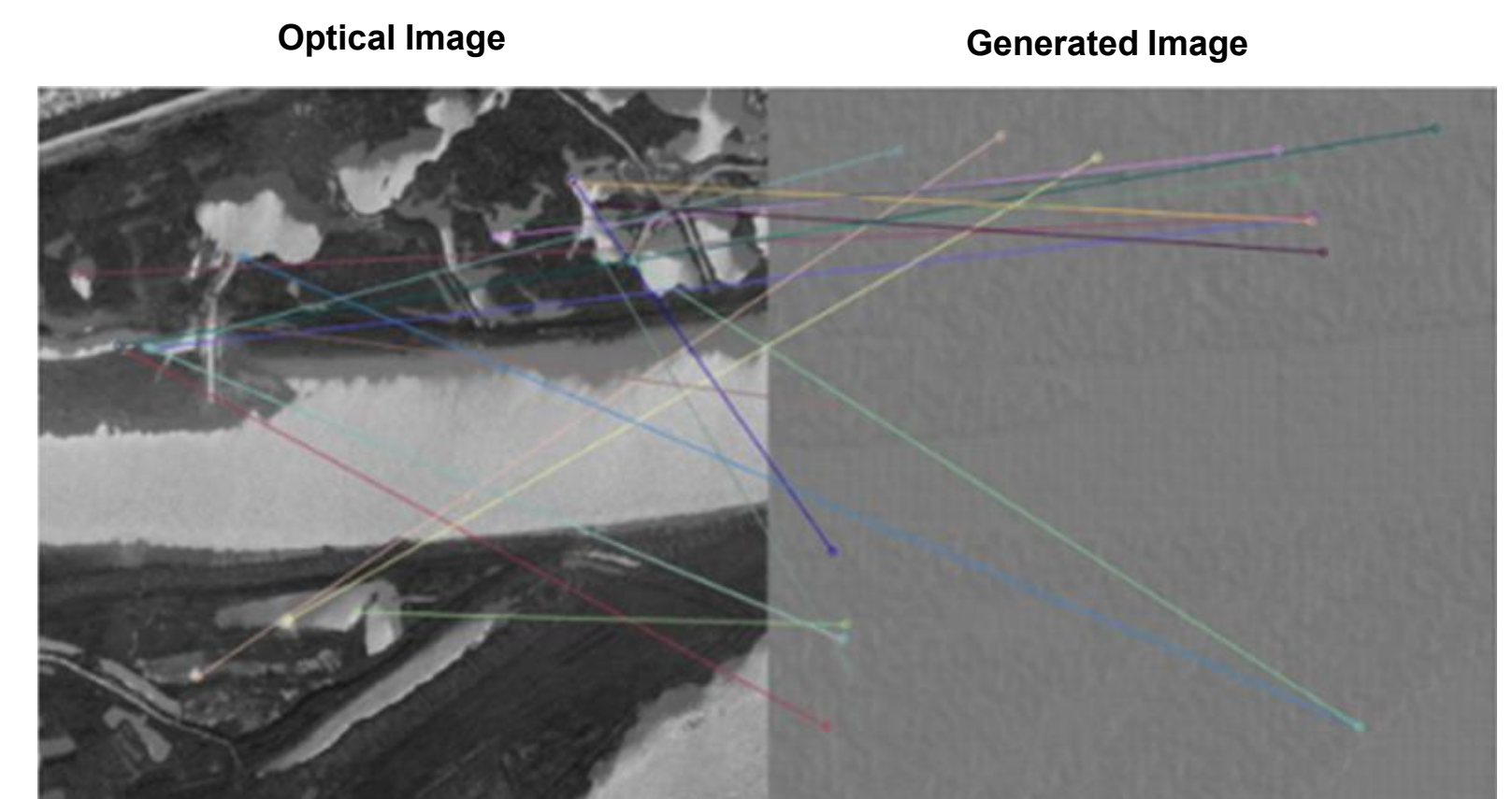


Figure 10: Feature based: SIFT and RANSAC

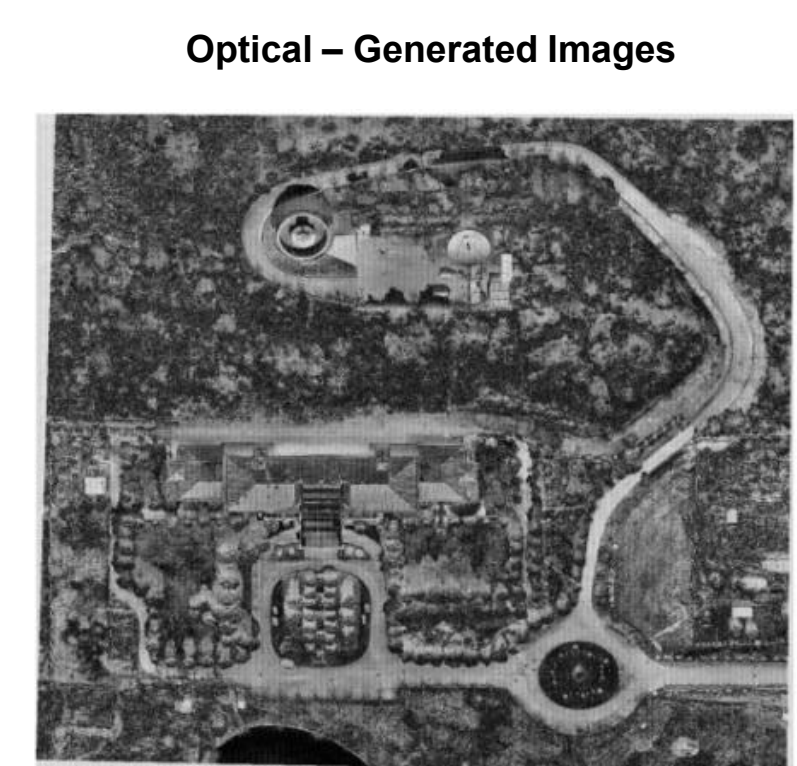


Figure 11: Pixel-wise Absolute Difference

| Method                           | Evaluation                                |
|----------------------------------|-------------------------------------------|
| Shape and contour-based matching | Matching score 0.94                       |
| Feature based: SIFT and RANSAC   | Structural similarity indices (SSIM) 0.20 |
| Pixel-wise Absolute Difference   | Score 0.81                                |

Table 2: Evaluation of Optical-Generated image matching

## Discussion

The results demonstrate significant improvements in image similarity metrics when using GAN-generated optical images compared to direct SAR-optical matching. The enhanced SSIM scores indicate better structural preservation and visual quality.

### Key Findings

- cGANs significantly enhance SAR-optical image registration accuracy
- Generated pseudo-optical images improve feature matching reliability
- UNet with skip connections effectively preserves spatial details
- Robust framework enables better multi-sensor data fusion

## Conclusions

Our findings demonstrate that cGANs significantly improve the matching of SAR and optical images by generating optical-like images from SAR data, as evidenced by higher SSIM and pixel-wise absolute difference scores (0.20 and 0.81, respectively). This suggests that cGANs can better preserve spatial and textural features, leading to more accurate keypoint matching and alignment with real optical images. These results are consistent with previous studies showing the effectiveness of deep learning models for SAR-to-optical translation and multi-sensor data fusion.

The practical implications are notable for remote sensing, as deep learning-based approaches offer flexible and scalable solutions for fusing data from different sensors and improving generalization across datasets. However, challenges remain, such as variability in image quality and the realism of generated images. Future work should investigate advanced generative models like CycleGANs or Transformers, and incorporate contextual information such as semantic segmentation to further enhance matching accuracy.

In this study, geometric differences between SAR and optical images were not addressed, but domain transformation using GANs shows strong potential for transferring spectral characteristics and supporting image registration, classification, and monitoring in rural and agricultural contexts. Continued research on geometry adjustments and spectral transformations is needed to address more complex scenarios in multi-sensor fusion.

## References

- [1] L. Wald, "Quality of high resolution synthesised images: Is there a simple criterion?" in Third Conference "Fusion of Earth Data: Merging Point Measurements, Raster Maps and Remotely Sensed Images", Sophia Antipolis, France, 2000, pp. 99–103, (hal-00395027).
- [2] S. Mahyoub, A. Fadil, E. M. Mansour, H. Rhinane, and F. Al Nahmi, "Fusing of optical and synthetic aperture radar (sar) remote sensing data: A systematic literature review (slr)," in International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences- ISPRS Archives, vol. 42, no. 4/W/12, 2019, pp. 127–138.
- [3] O. Somervold, M. Gazzea, and R. Arghandeh, "A survey on sar and optical satellite image registration," Remote Sensing, vol. 15, no. 3, 2023.
- [4] M. Schmitt and X. X. Zhu, "Data fusion and remote sensing: An ever-growing relationship," IEEE Geoscience and Remote Sensing Magazine, vol. 4, no. 4, pp. 6–23, 2016.
- [5] N. Merkle, S. Auer, R. Muller, and P. Reinartz, "Exploring the potential of conditional adversarial networks for optical and sar image matching," IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 11, no. 6, pp. 1811–1820, 2018.
- [6] M. Schmitt, L. H. Hughes, and X. X. Zhu, "The sent-2 dataset for deep learning in sar-optical data fusion," in ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, vol. 4, no. 1, 2018, pp. 141–146.
- [7] M. Huang et al., "The qxs-saropt dataset for deep learning in sar-optical data fusion," arXiv, 2021. [Online]. Available: <https://doi.org/10.48550/arxiv.2103.08259>