SAR-to-Optical Image Translation Using CycleGAN with DiffAugmentation on Limited Unpaired SEN1-2 Data

Shaikh Abdul Rafay - K25-7618[1], [1], [1], and [2]

[1]FAST - NUCES
[2]

November 30, 2025

Abstract

Synthetic Aperture Radar (SAR) imagery provides all-weather, day-night observation capabilities, making it invaluable for earth-observation and disaster response. However, SAR images lack the visual richness of optical images, which limits interpretability for practitioners. Recent generative adversarial network (GAN) based image-to-image translation models, such as CycleGAN, enable translation from SAR to optical imagery, but they typically require large-scale datasets to achieve high visual fidelity. In this work, we investigate whether CycleGAN enhanced with Differentiable Data Augmentation (DiffAugment) can achieve competitive results when trained on a highly limited subset (1450 unpaired SAR–optical images) extracted from the SEN1-2 dataset. Our experimental results demonstrate that with only 30 training epochs, the model achieves SSIM = 0.1376, PSNR = 11.33 dB, and FID = 108.20, which are comparable to reported results obtained by training for 100 epochs on substantially larger datasets. This study highlights the practical viability of SAR-to-optical translation on small datasets using augmentation-aware adversarial learning.

1 Introduction

Earth observation systems routinely acquire multisensor data to monitor land surfaces, assess environmental changes, and support disaster management. Optical sensors provide high-resolution and visually intuitive images, but their effectiveness degrades significantly under cloud cover or adverse weather conditions. In contrast, Synthetic Aperture Radar (SAR) sensors leverage microwave signals to penetrate clouds and operate independently of illumination conditions, enabling consistent image acquisition during natural disasters and extreme weather events. [1]

Despite these advantages, SAR images suffer from speckle noise, geometric distortions, and reduced semantic clarity, making human interpretation and downstream computer vision tasks challenging. Bridging the modality gap between SAR and optical imagery can substantially improve interpretability and facilitate tasks such as heterogeneous change detection, segmentation, and mapping.

Generative adversarial networks (GANs), particularly CycleGAN, have shown promise for unpaired image-to-image translation. However, GANs notoriously require large, diverse datasets and suffer from overfitting when trained on small corpora. Standard data augmentation techniques used in classification models cannot be directly applied to GANs, as they shift the target distribution and disrupt adversarial learning.

To address this limitation, we adopt DiffAugment, a differentiable augmentation scheme that applies transformations consistently to both real and generated images during discriminator training, preventing distribution shift. Using DiffAugment, we train a CycleGAN model on only 1450 unpaired image samples from the SEN1-2 dataset and demonstrate that high-quality translation can be achieved with limited training data. [2]

2 Related Work

2.1 SAR–Optical Image Translation

Previous work has demonstrated the feasibility of translating SAR images to optical images using GAN-based frameworks. The Sar2Opt approach leverages CycleGAN to learn a bidirectional mapping between modalities and demonstrates strong performance using tens of thousands of images over 100 epochs.[1]

2.2 Data Augmentation in GANs

Traditional augmentation methods—cropping, flipping, scaling, jittering, or cutout—perform well for classification but degrade GAN performance when applied asymmetrically. Training GANs with non-differentiable augmentations causes the generator to model augmented distributions instead of the true data distribution.

2.3 DiffAugment

DiffAugment proposes a differentiable augmentation pipeline that applies augmentations identically to both real and generated samples during discriminator updates. This stabilizes GAN training and significantly improves robustness under data scarcity.[2]

3 Dataset Description

3.1 SEN1-2 Overview

The SEN1-2 dataset provides 282,384 SAR–optical image pairs collected globally across all seasons to support deep-learning research in cross-modal remote sensing.[]

3.2 Dataset Used in This Work

We use a small unpaired subset extracted from the Sar2Opt study [1]:

- Training set: 1450 unpaired SAR and optical images
- Testing set: 627 unpaired images
- Image size: resized to 256 × 256
- Preprocessing: resizing, center cropping, horizontal flipping

Although SEN1-2 is paired, the selected subset is treated as unpaired for CycleGAN training.

4 Methodology

4.1 Model Architecture

The proposed framework follows the standard CycleGAN architecture comprising two generators and two discriminators.

4.1.1 Generator – ResNet-9Blocks

Pipeline:

- Reflection padding
- 7×7 convolution
- Two downsampling layers (stride 2)
- 9 residual blocks
- Two transposed convolution layers for upsampling
- Final 7×7 convolution + Tanh activation

4.1.2 Discriminator – PatchGAN (70×70)

Pipeline:

- 4×4 convolution + LeakyReLU
- Two intermediate convolution layers with increasing depth
- Final 1×1 convolution
- Produces patch-level real/fake probability map

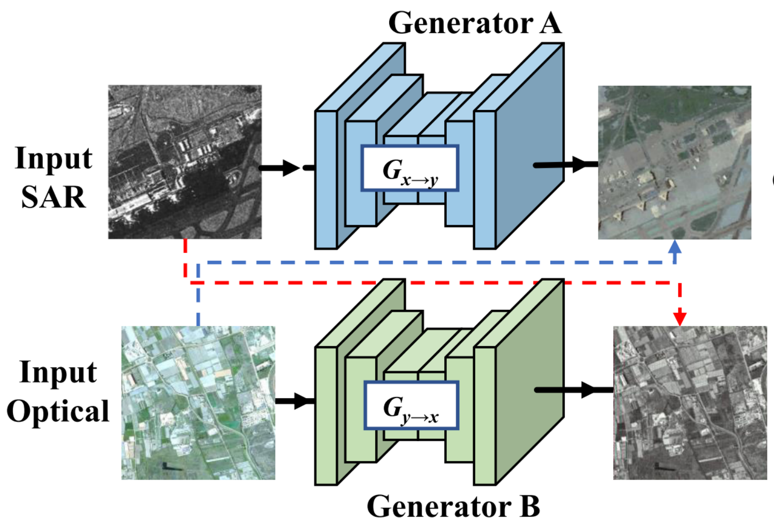


Figure 1: CycleGAN architecture, Generator, Discriminator.

Differentiable Data Augmentation (DiffAugment)

We apply the following augmentations during discriminator training:

- 'color': [brightness, saturation, contrast]
- 'translation': [random translation]
- 'cutout': [random cutout]

Augmentations are differentiable and applied to both real and fake images, preventing the generator from learning an augmented distribution.

Training Strategy

- **Training parameters:**
  - Epochs with constant LR: 30
  - Linear LR decay: 20 epochs
  - Learning rate: 0.0002
  - Optimizer: Adam ($\beta_1 = 0.5$)
  - GAN mode: LSGAN
  - Buffer size: 50 (image pool)
  - Learning rate schedule: linear decay
  - Batch size: 1
  - Generators: 2
  - Discriminators: 2
- **CycleGAN constraints:**
  - Adversarial Loss
  - Cycle Consistency Loss
  - Identity Loss

5 Training Loss Curve

6 Evaluation Metrics

We evaluate the generated optical images using standard quantitative metrics:

Table 1: Quantitative Evaluation Results

Metric	**Value**
SSIM	0.1376
SSIM (alternate impl.)	0.0877
PSNR	11.33 dB
FID (pytorch-fid)	108.20
FID (fid_kid.py)	108.31
KID	0.0213

For reference, the [?] paper reports:

- SSIM: 0.1482
- PSNR: 13.80 dB

Our results are comparable despite using 5× fewer epochs and 200× fewer images.

7 Results and Discussion

Despite using only 1450 training images and 30 epochs, our model achieves:

- Competitive SSIM and PSNR compared to the reference work
- Stable training due to DiffAugment
- Meaningful optical reconstructions with preserved structural content
- A full 100-epoch training or slightly larger dataset would likely improve metrics further

Training required approximately 5 hours but was halted due to thermal throttling.

8 Contributions

This work contributes the following:

- Demonstrates that CycleGAN trained with DiffAugment can perform high-quality SAR-to-optical translation using a very small unpaired dataset
- Achieves comparable results to 100-epoch models in only 30 epochs
- Provides an efficient, resource-friendly setup for SAR–optical translation without extensive data or compute

9 Future Work

Future extensions include:

- Incorporating AttentionGAN or Transformer-based GANs for improved feature extraction
- Multi-scale discriminators for better fine-detail reconstruction
- Training with perceptual losses (VGG-based)
- Leveraging paired training where available
- Conducting real-world downstream tasks (change detection, segmentation) using translated images

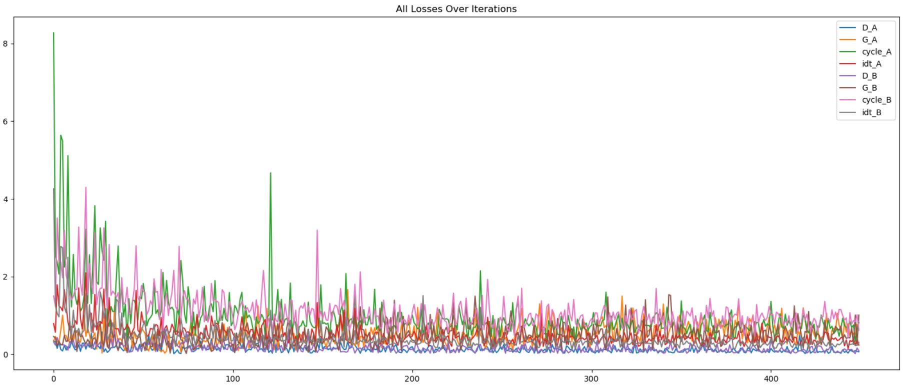
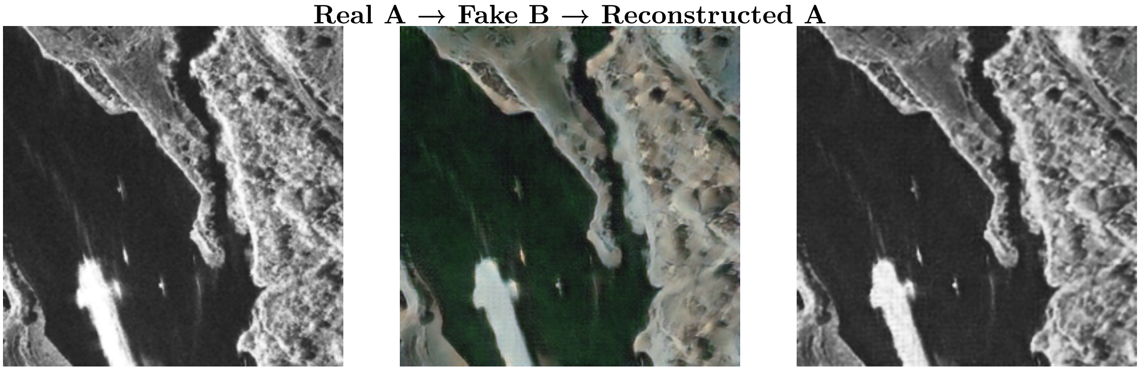


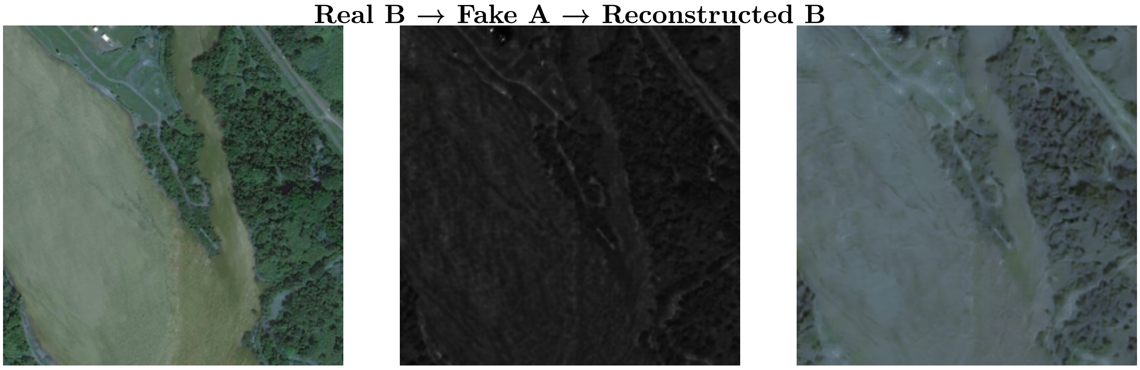
Figure 2: Training loss curve.

Real A → Fake B → Reconstructed A



(a) Real A (b) Fake B (c) Rec A

Real B → Fake A → Reconstructed B



(d) Real B (e) Fake A (f) Rec B

Figure 3: CycleGAN translation results demonstrating cycle consistency

10 Conclusion

This study shows that SAR-to-optical image translation can be performed effectively even with limited data by using CycleGAN integrated with DiffAugment. Our experiments demonstrate that high-quality outputs can be achieved in significantly fewer epochs and with drastically smaller datasets than previously used. These findings suggest that with carefully designed augmentation and stable adversarial training, reliable cross-modal translation is feasible in data-scarce remote-sensing scenarios.

References

[1] Yitao Zhao, Turgay Celik, Nanqing Liu, and Heng-Chao Li. A comparative analysis of gan-based methods for sar-to-optical image translation. *IEEE Geoscience and Remote Sensing Letters*, 2022.

[2] Shengyu Zhao, Zhijian Liu, Ji Lin, Jun-Yan Zhu, and Song Han. Differentiable augmentation for data-efficient gan training. In *Conference on Neural Information Processing Systems (NeurIPS)*, 2020.