NiR Channel Generation from RGB Satellite Imagery

Anonymous Author(s)
Affiliation
Address
email

1 Introduction

This study presents a framework for generating Near-Infrared (NIR) imagery from RGB satellite images using advanced generative models. We conducted a comprehensive analysis of the dataset, including spectral band correlations and spatial resolution challenges inherent to satellite data. A benchmarking review of existing solutions highlighted limitations in spectral fidelity and generalization for multi-modal remote sensing tasks. To address these gaps, we implemented and trained three architectures: Wasserstein GAN with Gradient Penalty (WGAN-GP), Discrete Flow Matching (DFM), and Continuous Flow Matching (CFM). Each model was optimized to preserve critical spectral features while translating RGB inputs to NIR outputs. Quantitative evaluation metrics (e.g., PSNR, SSIM) and qualitative assessments revealed distinct performance trade-offs: WGAN-GP demonstrated robustness in high-contrast scenarios, DFM achieved computational efficiency with minimal quality loss, and CFM excelled in reconstructing fine-grained spectral details. Ablation studies further analyzed the impact of training protocols and loss functions on output stability. The results underscore the potential of hybrid approaches combining adversarial training with flow-based dynamics for remote sensing applications, offering a pathway to reduce reliance on costly multi-spectral sensors. This work contributes to the field by providing a comparative framework for NIR synthesis, with implications for environmental monitoring, agriculture, and climate modeling.

2 Architectures

2.1 Wasserstein GAN with Gradient Penalty (WGAN-GP)

WGAN-GP stabilizes GAN training by minimizing the Wasserstein distance with a gradient penalty term, replacing weight clipping to enforce Lipschitz continuity on the critic. This avoids training artifacts and enhances generation quality, excelling in high-dimensional data synthesis.

2.2 Continuous Flow Matching (CFM)

CFM transforms noise into data via continuous ODE-based trajectories, preserving information and avoiding discretization errors. It combines the flexibility of diffusion models with flow efficiency, achieving state-of-the-art results in image and molecular generation (ref).

2.3 Discrete Flow Matching (DFM)

DFM approximates CFM through discrete steps, reducing computational costs while retaining theoretical benefits. Adaptive discretization (ref) enables scalable use in resource-limited tasks like audio synthesis or NLP.

Submitted to 38th Conference on Neural Information Processing Systems (NeurIPS 2024). Do not distribute.

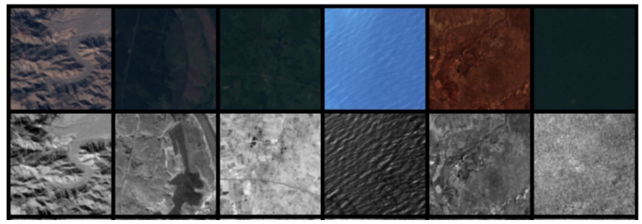

Figure 1: Sample figure caption.

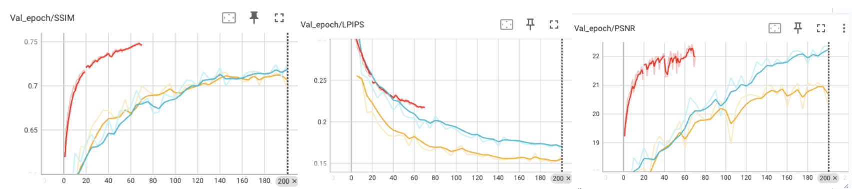

Figure 2: Sample figure caption.

2.4 Training

Training graphs illustrate the training dynamics of several models, monitored via SSIM, LPIPS, and PSNR metrics over 200 epochs. WGAN exhibits rapid improvement across all metrics during the initial stages, achieving high SSIM values (approximately 0.75) and PSNR (approximately 22.5 dB), along with a low LPIPS (approximately 0.22), by approximately 60-80 epochs. This observation aligns with the premise that the WGAN model trains more rapidly. Two other models (depicted in light blue and orange) demonstrate more gradual yet consistent improvement throughout the 200 epochs. By the culmination of training, the model represented by the light blue curve attains an SSIM of approximately 0.72, an LPIPS of around 0.17, and a PSNR of about 22.2 dB. The model indicated by the orange curve shows similar trends, reaching an SSIM of approximately 0.715, an LPIPS of circa 0.155 (the best among all models by the end of training), and a PSNR of approximately 20.8 dB. Overall, the graphs depict a protracted training process for most configurations, while concurrently confirming that WGAN converges considerably faster. It should be noted that the datasets employed for training are accessible on the GitHub platform.

3 Results and Analysis

	WGAN	DFM	CFM
SSIM (Higher is better, range [0,1])	0.7402	0.7002	0.7249
PSNR in dB (Higher is better)	22.1696	20.3391	22.4407
LPIPS (AlexNet) (Lower is better)	0.2263	0.1583	0.1659
SAM (Lower is better)	0.1611	0.1837	0.161
Correlation Coefficient (Higher is better, range [-1,1])	0.8385	0.7963	0.8432

Figure 3: Metrics results of trained models

Three distinct deep learning models—WGAN, DFM, and CFM—were evaluated for the task of translating RGB images to the Near-Infrared (NIR) spectrum. All three models demonstrated commendable performance across a suite of standard image quality metrics. Specifically, the WGAN model achieved the highest score for the Structural Similarity Index Measure (SSIM = 0.7402), whereas the DFM model attained the best result for the Learned Perceptual Image Patch Similarity (LPIPS = 0.1583), which reflects perceptual similarity. The CFM model emerged as the most balanced, exhibiting superior performance in Peak Signal-to-Noise Ratio (PSNR = 22.4407 dB), Spectral Angle Mapper (SAM = 0.1610 rad), and Correlation Coefficient (0.8432), alongside competitive SSIM and LPIPS values. It is pertinent to note that, based on a subjective assessment of training speed, WGAN was observed to be the most rapid to train. Considering the high performance levels achieved, it is postulated that the models' potential has not yet been fully realized, and continued training, coupled with hyperparameter optimization, could yield further enhancements in the results.

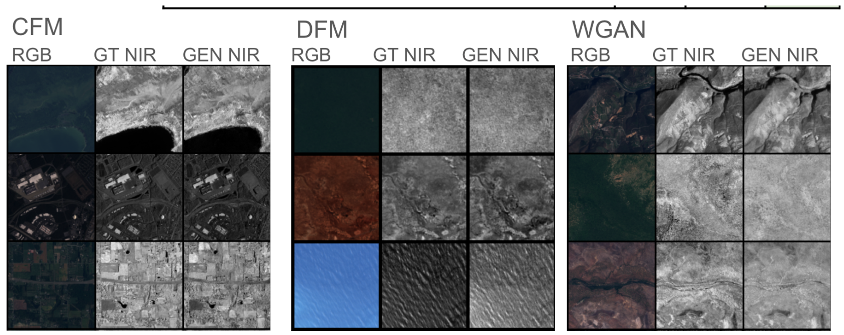

Figure 4: examples of generated images

4 Potential for Improvement

The strong results across the board are promising. It's plausible that the performance of these models has not yet reached a plateau. Continued training, potentially with fine-tuning of hyperparameters, learning rate schedules, could lead to further improvements in all metrics for all models.

References