Generative Models for Spectral Reconstruction from RGB Images

Li Luo

1. INTRODUCTION

Hyperspectral imaging (HSI) captures the spectral profile of real-world scenes across multiple wavelength bands, enabling the detection of unique spectral signatures for individual objects. This capability surpasses traditional RGB imaging by offering significantly more spectral detail, making HSIs invaluable for applications such as medical imaging, remote sensing, anomaly detection, autonomous driving, and more.

Despite their advantages, the acquisition of hyperspectral images is both expensive and technically challenging. The need for specialized equipment, such as advanced sensors and scanning mechanisms, limits their adoption in dynamic and real-time applications. In contrast, RGB imaging is widely accessible and cost-effective, driving interest in computational approaches to extract spectral information from RGB images—a process known as spectral reconstruction.

Spectral reconstruction (SR) from RGB images has been addressed using two primary approaches: prior-based methods and data-driven techniques. Prior-based methods leverage domain knowledge and statistical properties to constrain the reconstruction process. For example, sparse coding represents hyperspectral data as combinations of a few basis spectra, while manifold learning explores the low-dimensional structure inherent in hyperspectral images for accurate recovery. Gaussian process-based methods model spectral curves statistically, incorporating spatial and spectral priors, and spatially constrained approaches use spatial relationships to enhance local consistency in reconstructed images. These methods excel in scenarios with limited data but are constrained by their reliance on handcrafted priors.

In contrast, data-driven methods employ machine learning to map RGB images to hyperspectral representations. Convolutional Neural Networks (CNNs) were among the first models used, with advancements like residual and dense layers improving spectral fidelity. U-Net architectures further enhanced performance by combining encoder-decoder designs with skip connections to retain spatial details. Attention-based models, such as SRAWAN, focus on selectively emphasizing critical spatial and spectral features, improving reconstruction accuracy. The adaptive weighted attention network (AWAN) for SR achieve the AWAN investigate an adaptive weighted channel attention (AWCA) module to reallocate channel-wise feature responses via integrating correlations between channels. Generative models, particularly Generative Adversarial Networks (GANs) and diffusion models, have recently shown significant promise.

Our contributions are two-fold. First, we explore the generative models and compare vanilla Denoising Diffusion Possibility model with Pixel-based GAN which condition with RGB images to reconstruction hyperspectral images with noise estimation. Second, we do generalization tests and compare generative models with state-of-the-art deterministic models which prove that generative models are better to predict new datasets.

2. RELATED WORK

While data-driven methods generally offer superior accuracy and adaptability, they require large datasets and are prone to overfitting with limited data. Prior-based methods, on the other hand, are robust in such scenarios but lack the generalization capability of learning-based approaches.

Generative Adversarial Networks (GANs) are widely applied in spectral reconstruction tasks due to their capacity to model complex spatial and spectral details. SRCGAN employs a U-Net-based generator and a PatchGAN discriminator, focusing on spatial context enhancement. By combining adversarial loss, SRCGAN effectively preserves global image structure and reduces artifacts. This approach generates high-quality spectral images but is often hindered by training instability, limiting its reliability in certain applications.

Recently, diffusion models have gained attention for their generative capabilities. Emerging models like HSR-Diff and ISPDiff employ a forward diffusion process to add noise to RGB inputs and a reverse process for spectral reconstruction. These models excel in incorporating prior information and have demonstrated efficacy in fusing high-resolution MSI and low-resolution HSI data.

However, for the spectral reconstruction (SR) task, the inverse process recovers an HSI starting from random noise with the same size at the HSI; therefore, the noise space is large for conventional diffusion-based methods, which increases the uncertainty in predicting x_{t-1} in the inverse process. Degradation-Aware Diffusion Spectral Reconstruction (DDSR) extends conventional diffusion models by addressing prediction uncertainty in the inverse process, a common challenge when reconstructing hyperspectral images (HSIs) from RGB inputs. Unlike previous methods, DDSR incorporates degradation-aware corrections to model the degradation process (e.g., spectral downsampling, noise, JPEG compression) between HSIs and RGB images. This correction enables DDSR to efficiently guide the reconstruction process and adapt to diverse

degradation patterns. However, this correction is under degradation assumption and cannot apply for general reconstruction applications. We explore spectral reconstruction directly from RGB images without any assumption.

3. METHODS

3.1 Generative Adversarial Networks (GANs)

GANs are widely used for spectral reconstruction due to their ability to model complex distributions. The generator learns to produce HSIs, while the discriminator ensures their fidelity by distinguishing between real and generated images. We get the idea from SRCGAN which takes Conditional GAN as the main framework. The paper explores the use of Conditional Generative Adversarial Networks (cGANs) for reconstructing hyperspectral images (HSIs) from RGB inputs, integrating adversarial and spatial context-aware frameworks to optimize spectral reconstruction.

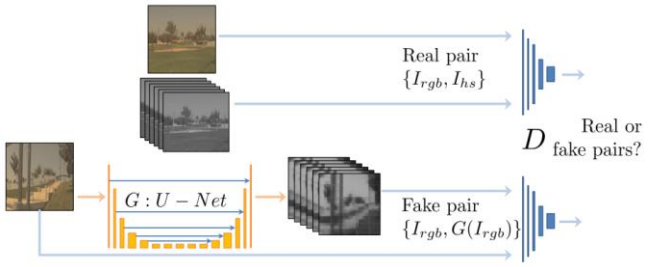

Figure 1. Adversarial spatial context-aware spectral image reconstruction model

GAN framework comprises: Generator (G): A neural network trained to produce realistic hyperspectral data. Discriminator (D): A classifier that distinguishes real HSIs from those generated by G. The two networks engage in a competitive learning process where G aims to deceive D while D improves its classification accuracy. This adversarial training minimizes the combined loss:

$$L_{\text{cGAN}}(G,D) = \mathbb{E}_{(x,y)\sim p_{\text{data}}}[\log D(x,y)] + \mathbb{E}_{x\sim p_{\text{data}},z\sim p_{\text{noise}}}[\log(1-D(x,G(x,z)))]$$

Here, x represents the RGB input, y the corresponding HSI, and zzz is the noise input.

For generator, A U-Net architecture enhances spectral reconstruction. Skip connections between encoding and decoding layers retain spatial detail, critical for preserving the structure of reconstructed HSIs. For discriminator, the PatchGAN discriminator focuses on local spectral consistency within patches of the reconstructed images. In training process, Loss functions combine adversarial objectives Ladv with L1 reconstruction loss to improve spatial coherence and global spectral accuracy:

$$L = L_{\text{adv}} + \lambda \ell_1$$

3.2 Diffusion Models

The diffusion process for spectral reconstruction consists of two phases: a forward process and a reverse process. During the forward process, noise is incrementally added to the hyperspectral image, simulating a degradation pattern as part of the training. In the reverse process, the model progressively denoises starting from random noise to reconstruct the hyperspectral image. This process is guided by the RGB image, which serves as a conditional input, providing partial spectral information to guide the reconstruction. The training process employs an L2 loss function, which minimizes the difference between the actual noise added during the forward process and the noise predicted by the model in the reverse process, conditioned on the RGB image. This setup ensures accurate and robust spectral reconstruction while leveraging the accessible information from the RGB input. Shown as Figure 2.

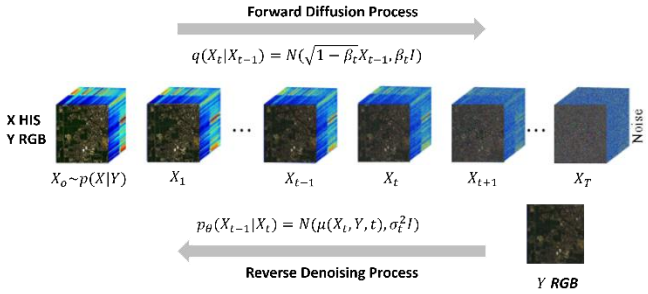

Figure 2. Hyperspectral reconstruction diffusion process

The denoising architecture we use U-Net as a backbone and consists of a number of ResBlocks and attention blocks. The U-Net consists of 3 encoder blocks and 3 corresponding decoder blocks, connected by skip connections. Each encoder block comprises a convolutional layer, batch normalization, and a Silu activation. The decoder blocks mirror the encoder, with transposed convolutions, batch normalization, and Silu activations. Additionally, skip connections are employed to transfer feature information across different levels of the network, ensuring the effective integration of both local and global image details. The network is designed to extract features at multiple scales, enriching the contextual information available for processing.

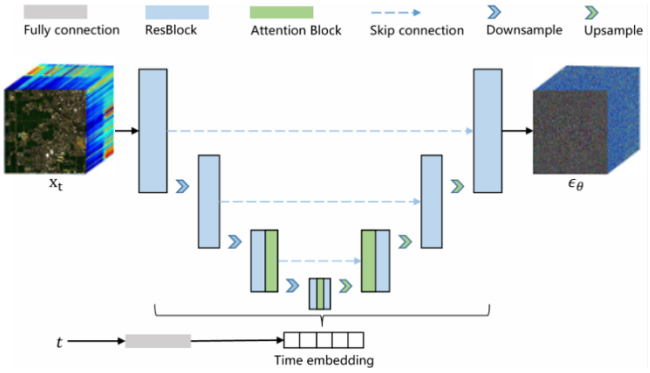

Figure 3. Denoising U-Net architecture

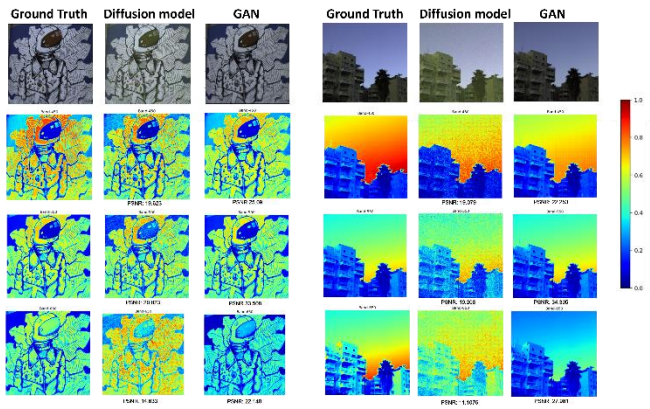

Figure 4. visual comparison between generative models

4. EXPERIMENTS

4.1 Implementation details

The GAN and diffusion models were trained using a learning rate of 0.0002 and a batch size of 4. The training was conducted for 1000 epochs on the ARAD-HS dataset, which contains 510 hyperspectral images with a spatial resolution of 512 x 482 and 31 spectral bands. ARAD-HS dataset is an HSI dataset for NTIRE-2020 with 510 images. The ARAD-HS dataset, with its diverse scenes and robust training splits, serves as a benchmark for assessing generative models. A large variety of indoor and outdoor scenes are collected, such as statues, vehicles and paintings. ARAD-HS splits 450 images for training, 30 images for validation and 30 images for testing.

The RMSE (Root Mean Square Error) metric was used to measure the reconstruction accuracy by comparing the generated hyperspectral images to the ground truth. The SAM (Spectral Angle Mapper) metric was also calculated to evaluate the spectral fidelity of the reconstructed images, where lower values indicate better alignment with the reference spectra. The quantitative results are shown in Table 1 and visual results are shown in Figure 3. We can see the vanilla conditional diffusion model perform poorly than SRCGAN, which need to be further optimized.

Table1. performance evaluation between generative models

Methods	RMSE	SAM
DDPM	0.15188	0.25722
SRCGAN	0.052613	0.044976

The diffusion model and GAN appear to generate images that reasonably match the ground truth data, though with some differences. This suggests these models can capture and reproduce key features of the underlying ground truth. The GAN outputs seem to capture some of the high-frequency spatial details present in the ground truth, while the diffusion model introduces noise and loses more color fidelity.

4.2 Generalization Test

To test the generalization for generative models, we use ICVL dataset which is also diverse natural hyperspectral image dataset. We randomly choose 5 representative scenes (windows, stone and tree are outdoor and color checker and stairs are indoors) and compare the quantitative results for each scene.

No additional preprocessing was applied to the ICVL dataset, and the same RMSE and SAM metrics were used to evaluate the performance of the GAN, diffusion, and state-of-the-art deterministic models (AWAN) on the selected ICVL scenes. This comparison allowed the researchers to assess how well the generative models can adapt to and perform on new, unseen hyperspectral data, beyond the training distribution.

Results show generative models have better overall performance, only the outdoor stone scene that AWAN and DDPM is very close in terms of RMSE.

Table2. prediction on ICVL outdoor scenes for generative models and AWAN network

Methods	window	stone	tree
AWAN	0.1484/0.2355	**0.2611**/0.3469	0.2168/0.4400
DDPM	0.2343/0.2	0.3405/0.2424	**0.2090**/**0.2331**
SRCGAN	**0.1307**/**0.1829**	0.3107/**0.2293**	0.2254/0.2872

Table3. prediction on ICVL indoor scenes for generative models and AWAN network

Methods	color checker	stairs
AWAN	0.5924/0.5269	0.24475/0.4042
DDPM	**0.4832**/**0.2745**	0.3328/**0.3102**
SRCGAN	0.5844/0.2913	**0.1984**/0.3761

5. CONCLUSION

Generative models, particularly GANs and diffusion models, offer promising pathways for spectral reconstruction from RGB images. Their ability to address the inherent challenges of this task makes them invaluable for advancing the field and enabling broader applications. To enhance generative models for spectral reconstruction: Further develop robust diffusion model to adjust loss functions focused on spectral accuracy as well as map to latent representation to mitigate the large noise space problem and better coverage for spectral dimensions. One the other hand, Design color correlation architectures that balance spatial and spectral feature extraction, integrating attention mechanisms and hierarchical structures.

REFERENCES

1. Zhang, Jingang, et al. "A survey on computational spectral reconstruction methods from RGB to hyperspectral imaging." Scientific reports 12.1 (2022): 11905.

2. Chen, Yunlai, and Xiaoyan Zhang. "DDSR: Degradation-Aware Diffusion Model for Spectral Reconstruction from RGB Images." Remote Sensing 16.15 (2024): 2692.

3. Saharia, Chitwan, et al. "Image super-resolution via iterative refinement." IEEE transactions on pattern analysis and machine intelligence 45.4 (2022): 4713-4726.

4. Alvarez-Gila, Aitor, Joost Van De Weijer, and Estibaliz Garrote. "Adversarial networks for spatial context-aware spectral image reconstruction from RGB." Proceedings of the IEEE international conference on computer vision workshops. 2017.

5. Shi, Shuaikai, Lijun Zhang, and Jie Chen. "Hyperspectral and multispectral image fusion using the conditional denoising diffusion probabilistic model." arXiv preprint arXiv:2307.03423 (2023).

6. Zeng, Haijin, et al. "Unmixing Diffusion for Self-Supervised Hyperspectral Image Denoising." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2024.

7. Li, Jiaojiao, et al. "Adaptive weighted attention network with camera spectral sensitivity prior for spectral reconstruction from RGB images." *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*. 2020.