Spectral image generation through generative adversarial networks using RGB images

Emmanuel David Martinez Estrada

Degree Project to qualify for the Master's degree in Systems Engineering and Informatics

Supervisor

Henry Arguello Fuentes

Ph.D. in Electrical and Computer Engineering

Universidad Industrial de Santander

Faculty of Physicomechanical Engineering

School of Systems Engineering and Informatics

Bucaramanga

2024

Contents

List of Figures

List of Tables

Abstract

Title: Spectral image generation through generative adversarial networks using RGB images[1]

Author: Emmanuel David Martínez Estrada[2]

Keywords: Spectral image generation, generative adversarial networks, RGB images, compressive spectral imaging.

Description: Spectral imaging involves acquiring and processing electromagnetic radiation stored as a spectral image. The spectral images are acquired through scanning or compressive spectral imaging (CSI) methods. Although spectral images are widely used in deep learning for spectral image reconstruction, classification, or anomaly detection, their acquisition is a challenging task due to physical and cost limitations. Therefore, this research proposes generating spectral images through a generative adversarial network (GAN) guided by RGB images. Specifically, a generator model is trained using a random noise input to generate the spectral images. The resulting spectral images are then processed using a spectral response function to obtain RGB-mapped images. Posteriorly, in an adversarial manner, a discriminator model is used to determine whether the generated RGB-mapped image is real or fake. An implicit regularization is also proposed to generate a realistic spectral image, where a pre-trained model with real spectral images is used during GAN training. Additionally, a spectral band post-processing normalization step is performed over the generated images. Finally, to validate the generated spectral images, a data augmentation strategy is performed on the CSI task where we find that a few real captured spectral images are enough to improve CSI performance.

---

[1] Degree work

[2] Faculty of Physicomechanical Engineering. School of Systems Engineering.
Supervisor: Henry Arguello Fuentes. Doctor in Electrical and Computer Engineering.

Resumen

Título: Generación de imágenes espectrales a través de redes adversarias antagónicas usando imágenes RGB[3]

Autor: Emmanuel David Martínez Estrada[4]

Palabras Clave: Generación de imágenes espectrales, redes generativas adversarias, imágenes RGB, imágenes espectrales compresivas.

Descripción: La obtención de imágenes espectrales consiste en capturar y procesar la radiación electromagnética que luego se almacena como una imagen. Se adquieren mediante métodos de escaneo o imágenes espectrales compresivas (CSI), y aunque su uso es amplio en aprendizaje profundo para tareas como reconstrucción, clasificación o detección de anomalías, la captura de estas imágenes es desafiante por sus limitaciones físicas y costos elevados. Esta investigación propone generar imágenes espectrales mediante redes generativas adversarias (GAN) utilizando imágenes RGB. Para ello, un modelo generador, entrenado con ruido aleatorio, produce imágenes espectrales. Estas imágenes se procesan con una función de respuesta espectral que las mapea en RGB, y luego un modelo discriminador evalúa si la imagen mapeada es real o falsa. Se propone además una regularización implícita que emplea un modelo preentrenado con imágenes espectrales reales para asegurar la calidad de las imágenes generadas. Posteriormente, se aplica una normalización de banda espectral en el posprocesamiento. Finalmente, las imágenes generadas son validadas a través de una estrategia de aumento de datos en la tarea de CSI, donde se comprueba que unas pocas imágenes espectrales reales pueden mejorar el rendimiento del método CSI en la reconstrucción.

---

[3] Trabajo de Grado

[4] Facultad de Ingeniería Fisicomecánicas. Escuela de Ingeniería de Sistemas.
Director: Henry Arguello Fuentes. Doctor en Ingeniería Eléctrica y Computación.

Introduction

Spectral imaging refers to the technique of acquiring and analyzing electromagnetic radiance emitted by light sources to obtain spectral images as 3D data cubes. Spectral information enables the identification and analysis of the materials present due to their unique chemical compositions. Therefore, this technique is widely used in a variety of fields, such as remote sensing, medical imaging, and agriculture Shaw and Burke (2003); Salamí et al. (2014); Gu et al. (2021); Clancy et al. (2020); Genser et al. (2020). Typically, spectral images are acquired using scanning methods. These methods involve capturing one-dimensional or two-dimensional images and combining them to form the final spectral image Yuan et al. (2021). Nevertheless, this approach requires moving the optical imager, the scene, or the filters to capture specific spectra, making the process time-consuming and computationally expensive.

To overcome this limitation, compressive spectral imaging (CSI) has emerged as a methodology for the acquisition and reconstruction of spectral images, where optical systems are designed to capture a single 2D coded measurement from a scene, which is then reconstructed using computational algorithms Bacca et al. (2023); Yuan et al. (2021). However, these systems require complex setup and calibration processes, which can be challenging to implement in real-world environments Hagen and Kudenov (2013). As a result, extensive spectral image datasets do not exist, which are used for data-driven deep learning methods, such as spectral image reconstruction, and classification, among others Wang et al. (2019); Meng et al. (2020); Li et al. (2019); Hong et al. (2020).

To address this issue, various algorithms have been proposed to generate samples for spectral image classification tasks Liu et al. (2020b), where spectral signatures or small spatio-spectral patches are generated. However, the generation of entire spectral images has not been extensively explored yet. For instance, SHS-GAN Hauser et al. (2021) employs an RGB and spectral image dataset to perform unsupervised spectral reconstruction from the RGB images and discriminate them using the spectral dataset. Similarly, Zhu et al. Zhu et al. (2021) introduces an unsupervised method for reconstructing spectral images from RGB images by leveraging unpaired datasets and semantic information. However, while these models aim to obtain novel spectral images, they primarily perform a mapping process using unpaired datasets. In contrast to the mentioned methods, LD-GAN Martinez et al. (2023) generates novel spectral images from random noise following a low-dimensional (LD) representation employing an autoencoder network. Similarly, the current research has a similar approach where spectral images can be generated from a random input noise, but different from LD-GAN, the spectral image generation will be guided by RGB images.

Unlike spectral image datasets, RGB image datasets are abundant because of their straightforward acquisition through commonly used devices such as smartphones or professional cameras. These devices capture three specific electromagnetic wavelengths from a scene (R: red, G: green, and B: blue) Martínez et al. (2020a) forming the RGB images. Furthermore, the widespread use of social media platforms and the increasing adoption of deep learning for computer vision applications Qi et al. (2021); He et al. (2017); Liang et al. (2021) has led to the availability of numerous RGB images on the internet, which have been recompiled forming large RGB image datasets freely accesible Deng et al. (2009); Liu et al. (2015).

Therefore, this research proposes an algorithm to generate spectral images from random noise guided by RGB images. The proposed method considers using a generative adversarial network (GAN) Goodfellow et al. (2020), a spectral response function for obtaining mapped-RGB images, a proposed implicit regularization function to inject spectral information into generated spectral images, and a spectral band post-processing normalization step to improve the spectral information quality. Specifically, the generator model takes random noise inputs and produces spectral images, which are mapped to RGB images using the spectral response function. The mapped-RGB images are discriminated by a discriminator model to distinguish real from fake samples. To ensure that the generated spectral images contain realistic spectral information, the implicit regularization function is proposed, which indirectly injects spectral information using a pre-trained model with spectral images. To enhance the spectral information of generated images, a spectral band post-processing normalization step is applied to adjust the mean and the standard deviation of each spectral channel. Finally, to validate the proposed method, the generated spectral images will be employed as a data augmentation strategy for the CSI task using perceptual and quantitative metrics such as peak signal-to-noise ratio, structural similarity index, and spectral angle mapper, which show that the generated spectral images in this research enables accurate spectral image reconstruction.

1. Objectives

General Objective

- To design an algorithm for the generation of spectral images through generative adversarial networks based on RGB images for being used in snapshot spectral imaging.

Specific objectives

- To search and select a large dataset of RGB images for generating new spectral images.
- To develop a mathematical model and an optimization algorithm based on generative adversarial networks for the generation of spectral images from RGB images considering a spectral response function.
- To implement the generative and discriminative model with adversarial training based on the developed optimization algorithm in order to generate new samples from the selected dataset.
- To evaluate and validate the designed algorithm for spectral image generation by recovering spectral images from 2D encoded measurements through simulations.

2. State-of-the-art

2.1. Spectral Imaging

Spectral imaging is a group of techniques to collect the emitted electromagnetic radiation from objects and post-processing until obtaining a 3D discrete spectral data cube representation known as a spectral image. A spectral image can be represented as $\mathscr{X} \in \mathbb{R}^{M\times N\times L}$, where (M,N) are the discrete spatial coordinates, and L is the number of spectral bands. A straightforward representation is given by a matrix representation $\boldsymbol{X} \in \mathbb{R}^{L\times MN}$, where each column represents a spectral signature, as shown in Fig. 1. Given the nature of spectral signatures, it is possible to identify the contained materials, having a wide range of applications in remote sensing, medical imaging, environmental monitoring, agriculture, and anomaly detection, among others Gat et al. (1997); De Biasio et al. (2010); Li et al. (2018b); Azimi et al. (2018); Bartczak et al. (2018); Clancy et al. (2020). In this section, we explore spectral image acquisition based on scanning methods and snapshot spectral imaging with their respective advantages and drawbacks.

2.1.1. Spectral image acquisition.

The main methods for acquiring spectral images are divided into scanning techniques and compressive sensing, often implemented through snapshot spectral systems.

Scanning methods: The spectral image acquisition relies on scanning methods using 1D or 2D sensors, as spectral images contain 3D information, as shown in Fig. 2. Specifically, Whiskbroom scanners perform a pixel-by-pixel scan and obtain each spectral signature until the image

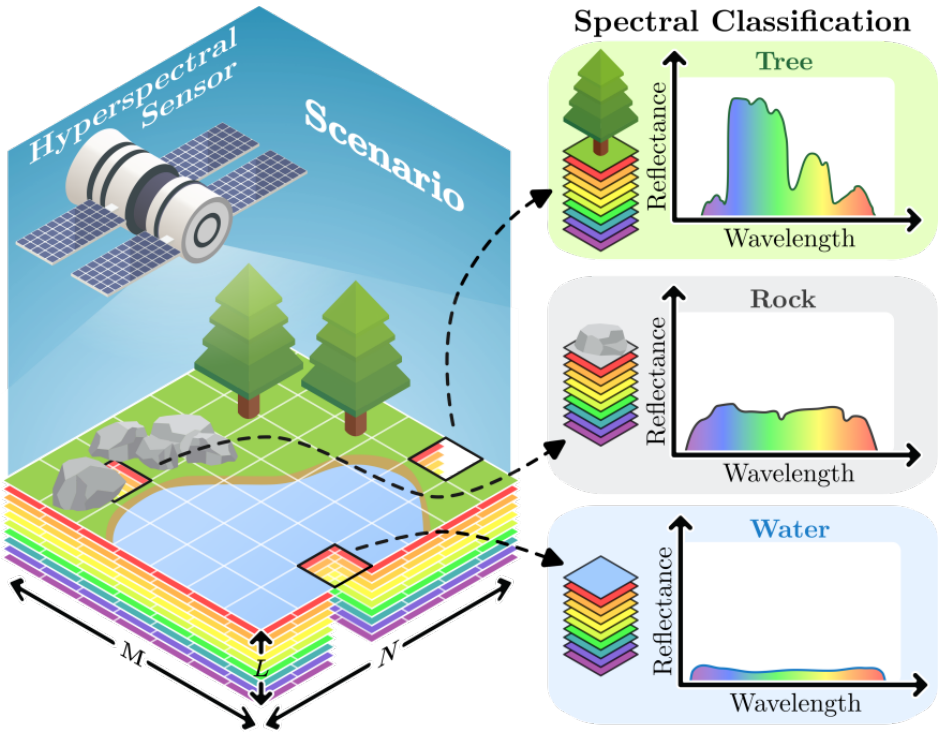

Figure 1

Spectral imaging and classification from three different spectral signatures. Source: Bacca et al. (2023)

is formed. Pushbroom scanners sweep over the scene line-per-line, forming the spectral image. Whiskbroom and pushbroom scanners are widely used in remote sensing, where these optical systems are installed in airborne or aircraft and take advantage of the motion of the earth's orbit or oneself Fowler (2014). Wavelength scanners are used in microscopy to capture specific wavelengths of fluorescent molecules for a more precise analysis of their compositions Zimmermann et al. (2003). As mentioned above, using 1D or 2D sensors is necessary because 3D or high-dimensional sensors generally do not exist. One drawback of scanning methods is that they can only capture static objects and rely on the precise calibration of the sensor to do so accurately.

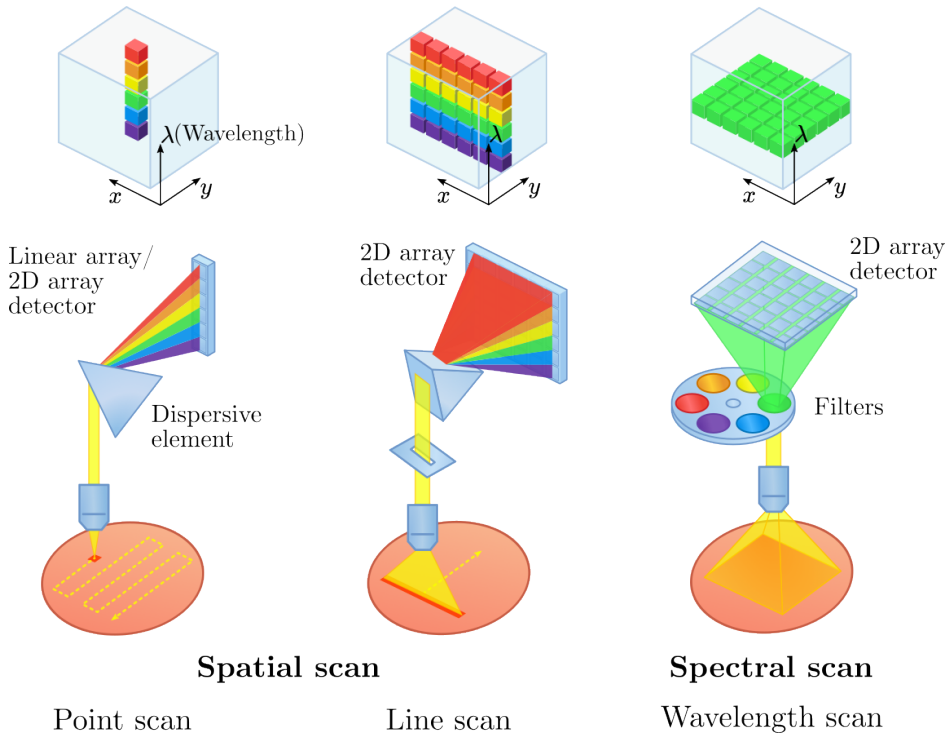

Figure 2

Spectral acquisition methods based on pixel, line, and wavelength scanning. Source: Wang et al. (2017)

Additionally, capturing high spatial-spectral resolution images involves a trade-off due to these technological limitations Chaudhuri and Kotwal (2013).

Compressive Spectral Imaging: The objective of compressive spectral imaging (CSI) is to capture a spectral image $\boldsymbol{x} \in \mathbb{R}^n$ as a 2D projected measurement. The acquired measurement can be represented as $\boldsymbol{y} = \boldsymbol{H}\boldsymbol{x}$, where $\boldsymbol{y} \in \mathbb{R}^m$ and $\boldsymbol{H} \in \mathbb{R}^{m \times n}$ represents the optical system as a measurement matrix that used to be a sparse matrix. Note that $\boldsymbol{y}$ is a compressed representation of $\boldsymbol{x}$, where $m \ll n$.

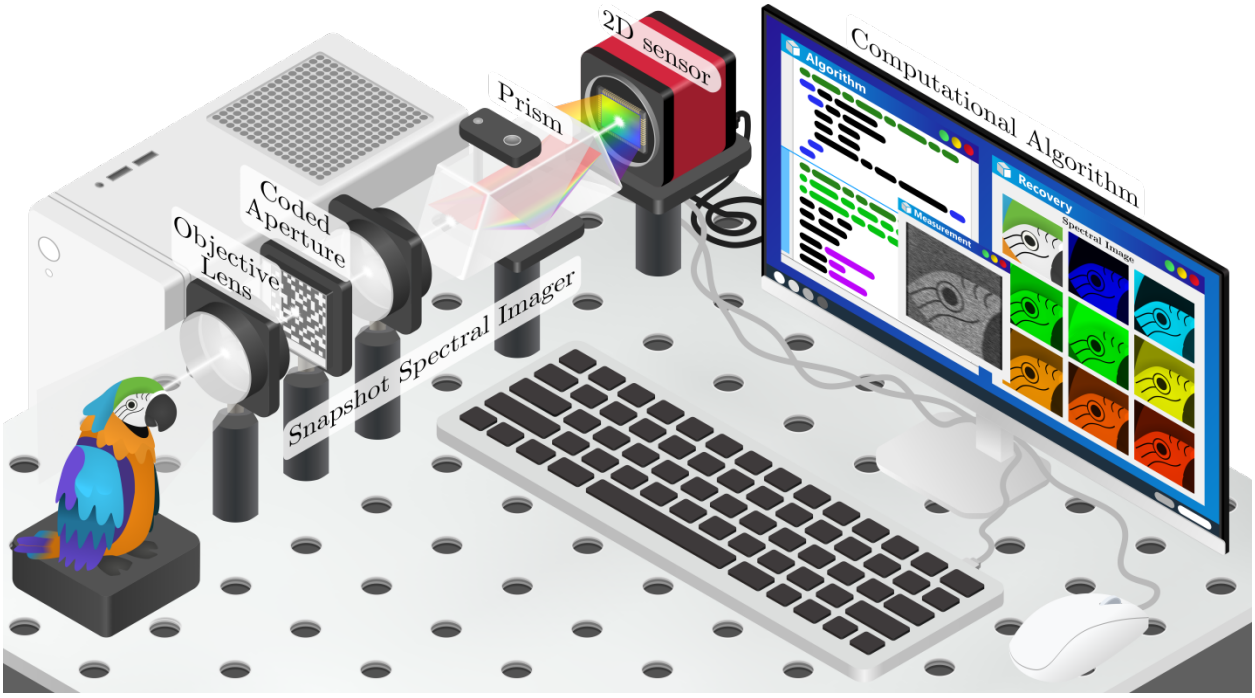

Figure 3

CASSI system. The emitted light by the scene passes through a coded aperture, and the prism disperses the coded light until it is integrated into the 2D sensor. Source: Bacca et al. (2023)

CSI optical systems code the emitted light by the scene through two main optical elements: coded apertures (CAs) and prisms. The CAs allow pass or block some spatial coordinates from light and prism disperse the light spectrum. The most known optical system is the coded aperture snapshot spectral system (CASSI) Arce et al. (2013), shown in Fig. 3. The optical path of the CASSI system codes the incoming light through a CA and posteriorly disperses the coded light until it is integrated into a 2D sensor generating the 2D coded measurement. There are other variations of the CASSI system, where other light elements and codification patterns are employed to enrich the capture, such as DD-CASSI, 3D-CASSI, and Color-CASSI systems Arguello and Arce (2014); Gehm et al. (2007); Cao et al. (2016).

In contrast to scanning methods, CSI optical systems can capture a scene in a single shot, enabling applications in video. However, both methodologies involve complex configuration and calibration of snapshot spectral systems, requiring specialized equipment such as the mentioned CAs and prisms. Additionally, the extensive spatial requirements of CSI optical systems make them impractical for use in uncontrolled environments, such as outdoor settings.

2.1.2. Spectral image reconstruction. To recover the desired spectral image from 2D coded measurements, the following methodologies have been widely explored in the state-of-the-art.

Handcrafted algorithms: 2D coded measurements are processed by computational algorithms that can consider prior information. The reconstruction of the spectral image is given by solving the optimization problem

$$\hat{\boldsymbol{x}} = \arg\min_{\boldsymbol{x}} \mathscr{L}(\boldsymbol{H}\boldsymbol{x}, \boldsymbol{y}) + \mathscr{R}(\boldsymbol{x}), \tag{1}$$

where $\mathscr{L}(\cdot,\cdot)$ is a loss function that approximates the 2D coded projection to the simulated matrix measurement with the desired spectral image $\boldsymbol{x}$, such as $||\cdot||_p$, for $p \in \{1,2,\infty\}$. $\mathscr{R}(\cdot)$ is a regularization function that contains prior information such as sparsity, low-rank, or self-similarity Gelvez and Arguello (2020); Li et al. (2020); Bacca et al. (2021); Gelvez-Barrera et al. (2022).

Deep learning algorithms: Given a dataset $\{\boldsymbol{x}^{(i)}, \boldsymbol{y}^{(i)}\}_{i=1}^{K}$ of K pairs of spectral images

$\boldsymbol{x}^{(i)} \in \mathbb{R}^{MNL}$, where (M,N) represents their spatial resolution and L is the number of spectral bands, with their respective measurements $\boldsymbol{y}^{(i)} = \boldsymbol{H}\boldsymbol{x}^{(i)}$. Through a deep neural network (DNN) $\mathscr{N}_{\boldsymbol{\theta}}$ with $\boldsymbol{\theta}$ parameters, the optimization problem can be expressed as

$$\boldsymbol{\theta} = \arg\min_{\boldsymbol{\theta}} \sum_{i=1}^{K} \mathscr{L}\left(\mathscr{N}_{\boldsymbol{\theta}}(\boldsymbol{y}^{(i)}), \boldsymbol{x}^{(i)}\right) + \mathscr{R}\left(\mathscr{N}_{\boldsymbol{\theta}}(\boldsymbol{y}^{(i)})\right), \tag{2}$$

where the objective is to optimize the $\boldsymbol{\theta}$ parameters of the DNN, and in this case, $\mathscr{R}(\cdot)$ regularizes the optimization function using the given dataset.

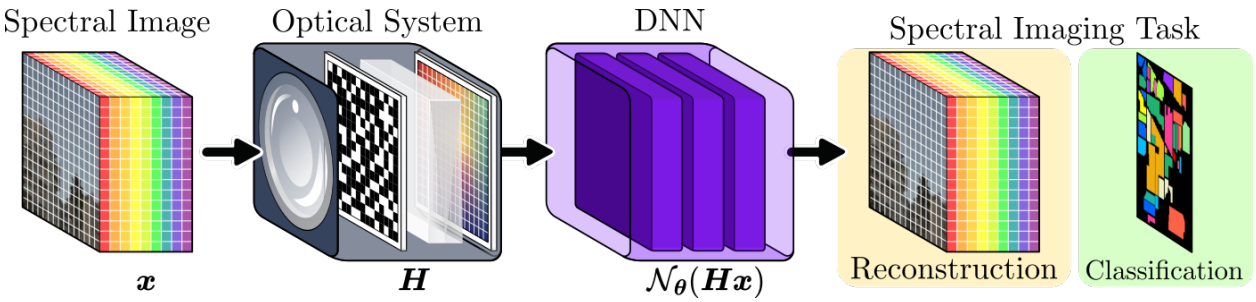

Figure 4

Deep learning-based Spectral image reconstruction or classification from 2D coded measurements.

The DNNs used for solving the reconstruction or classification tasks are based on popular architectures, such as Unet, GANs, and transformers Ronneberger et al. (2015); Goodfellow et al. (2020); Vaswani et al. (2017), as shown in Fig. 4. For spectral image reconstruction, iterative algorithms such as FISTA, TwIST, HQS and ADMM Beck and Teboulle (2009); Bioucas-Dias and Figueiredo (2007); Wang et al. (2019); Boyd et al. (2011) have been merged with DNNs to pro-

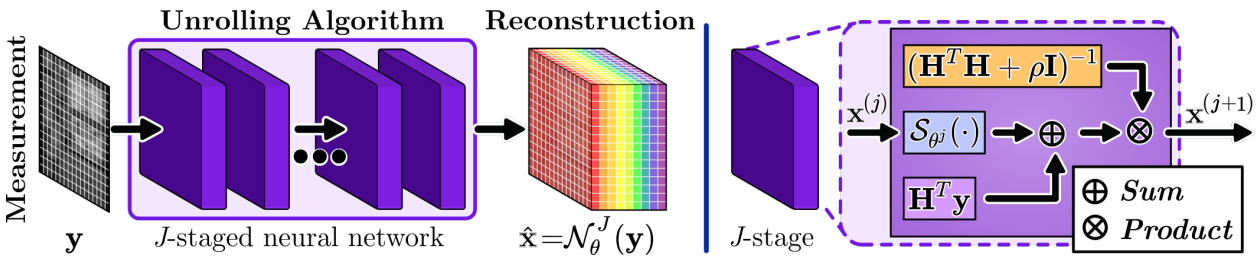

Figure 5

Unrolling algorithm. Spectral image reconstruction or classification from 2D coded measurements based on classic iterative algorithms.

vide solutions with stronger and interpretable results. Some methods suggest unrolling traditional optimization methods and learning the proximal operators resulting in an interpretable deep model Xiang et al. (2021); Sun et al. (2016); Wang et al. (2019), as shown in Fig. 5. Specifically, the network is composed for J *stages* $\hat{\boldsymbol{x}} = \mathscr{N}_{\boldsymbol{\theta}}(\mathbf{y}) = \mathscr{N}_{\boldsymbol{\theta}}^{J}\left(\mathscr{N}_{\boldsymbol{\theta}}^{J-1}\left(\cdots \mathscr{N}_{\boldsymbol{\theta}}^{1}\left(\mathbf{y}\right)\cdots\right)\right)$, of the form

$$\mathscr{N}_{\boldsymbol{\theta}}^{J}(\mathbf{y}) = (\boldsymbol{H}^T\boldsymbol{H}+\rho\boldsymbol{I})^{-1}\left(\boldsymbol{H}^T\boldsymbol{y}+\rho\mathscr{S}_{\boldsymbol{\theta}^J}\left(\boldsymbol{f}^{J-1}\right)\right), \tag{3}$$

which is inspired by a HQS formulation with $\mathscr{S}_{\boldsymbol{\theta}}(\cdot)$ as the proximal operator of a deep spectral prior Boyd et al. (2011); Wang et al. (2019). Figure 5 shows a usual representation of the unrolled-based scheme, where each feature (stage) layer represents a spectral image estimation. Most recent approaches propose a joint design of the optics and DNN. In this case, the measurement matrix $\boldsymbol{H}$ is employed as a trainable layer performing an end-to-end optimization.

2.2. RGB Imaging

RGB imaging is a group of techniques to capture three specific wavelengths, i.e, red, green, and blue, (RGB) from emitted light by an object and post-processing until obtaining an RGB

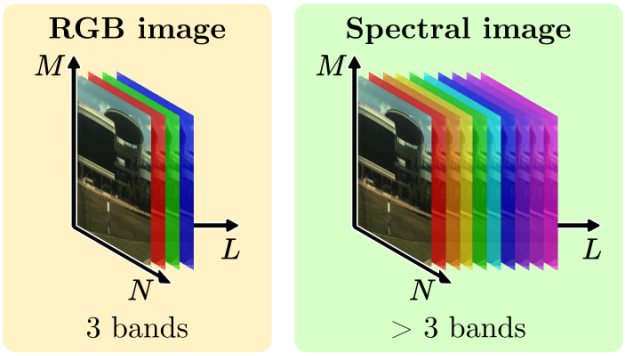

Figure 6

RGB image vs spectral image. Spectral images contain a wide range of bands different from RGB images which only contain three bands.

image, which discrete representation of the captured light with only three bands, as shown in the Fig. 6. In this case, an RGB image can be represented as $\mathscr{X} \in \mathbb{R}^{M \times N \times 3}$ and can be considered as a degraded spectral image. Due to the limited number of spectral bands in RGB images, the capability to identify materials, as achieved by spectral images, may not be feasible. However, these images are used for many applications related to image processing and computer vision, such as object detection or segmentation, image restoration, pattern recognition, and human action, among many more tasks He et al. (2017); Pak and Kim (2017); Li et al. (2018a); Mehta et al. (2018); Liang et al. (2021); Kong and Fu (2022). This section covers the acquisition of RGB images and their relevance in computer vision tasks.

2.2.1. RGB image acquisition. RGB image acquisition is given in cameras that use three color filters to capture the emitted light by the scene. The color filters are based on the natural spectral response of human vision, provided by three different types of photoreceptor cells in the eye Hunt (1995). Combining the acquired specific wavelength images forms the color

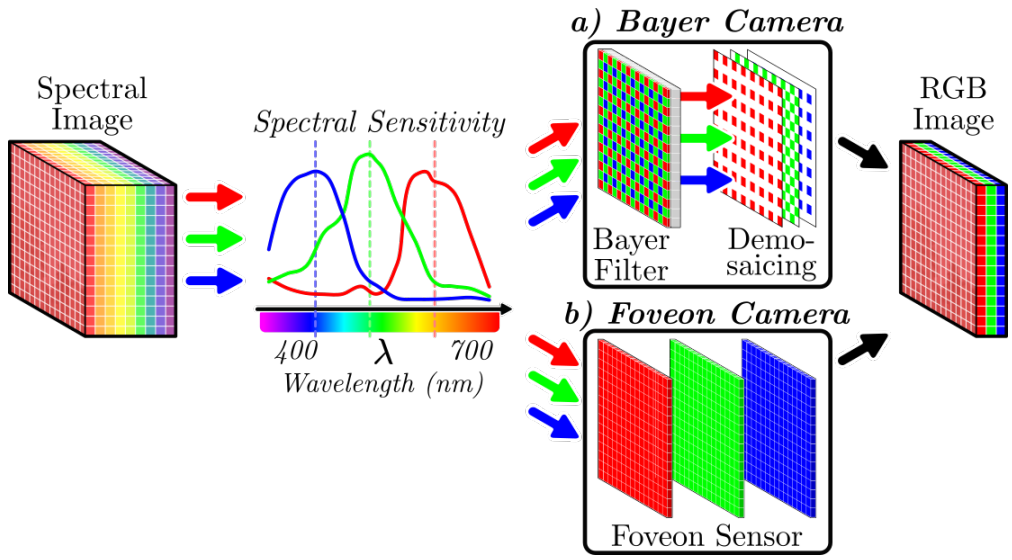

Figure 7

RGB image acquisition. Three specific wavelengths (RGB) are captured in two ways. a) By a Bayer pattern and a demosaicing algorithm, or b) By a Foveon camera that captures the complete RGB wavelengths.

or RGB image. Standard use of the color filters has a spatial distribution based on the Bayes pattern obtaining a raw image, which it is necessary to apply a demosaicing algorithm to obtain the desired RGB image Martínez et al. (2020b), as shown in Fig. 7 a). Assuming ideal atmospheric and illumination conditions, the RGB image acquisition can be approximated as a linear spectral degradation of the spectral image as $\boldsymbol{y} = \boldsymbol{R}\boldsymbol{x}$, where $\boldsymbol{R} \in \mathbb{R}^{MN3 \times MNL}$ represent the discrete spectral response function of the used 2D sensor and $\boldsymbol{y} \in \mathbb{R}^{MN3}$ is the vectorized form of the RGB image. It should be noted that this approximation is more similar to a Foveon camera, which does not require a pattern filter and directly captures each wavelength with individual sensors one on top of the other, as shown in Fig. 7 b). Unlike spectral images, it is straightforward to develop optical systems that allow the acquisition of RGB images, such us smartphones and professional cameras

Gastaldi et al. (2005).

2.2.2. RGB image computational tasks. RGB images have gained significant relevance for computer vision and deep learning applications, such as **Image classification**, where each image is labeled with a name that refers to its content Dosovitskiy et al. (2020); Li et al. (2022). **Object segmentation**, where each pixel in an image is labeled, allowing a better understanding of the objects in the image Caelles et al. (2017); Luiten et al. (2018). **Image colorization**, which adds colors to grayscale images. This is useful for old images, which used to be black and white Zhang et al. (2016); Su et al. (2020). **Image restoration**, which improves the quality of a damaged image. For example, denoising, blurring, or inpainting Liang et al. (2021); Wang et al. (2022). **Super-resolution**, which increases the spatial resolution of an image without loss of perceptual quality Dong et al. (2015); Zhang et al. (2018). **Style transfer**, which transfers the style from one image to another image without loss of consistency. For example, picture to photorealism image, zebra to the horse, etc. Gatys et al. (2015); Zhu et al. (2017). Due to the variety of applications where RGB images are used, exist large RGB image datasets for solving these tasks Deng et al. (2009); Liu et al. (2015), different from spectral images, which lack data due to their complex acquisition.

2.3. Generative Adversarial Networks

A challenging computational task in the computer vision community is generating realistic images from a given dataset. The main motivation is to optimize a model that samples new images following an unknown distribution from the given dataset. Most relevant techniques rely on generative adversarial networks. However, other methods have also been explored such as flow-based

models, energy models, and the most recent, diffusion models. In this research, we work with generative adversarial networks (GANs) because they have been widely explored in the state-of-the-art and are more suitable since generative models rely on a strong composition of Bayesian and statistical theory which is different for each model difficulting to develop a suitable model. However, an exploration of the proposed research could be developed for other generative models as future work.

Sampling consists of extracting a subset of samples $x \sim \mathcal{N}(\mu, \sigma^2)$ from a chosen distribution. In this case, a normal distribution with mean μ and standard deviation σ. Thus, estimators from extracted samples must approximate the estimators of the chosen distribution. Following the same idea, GANs Goodfellow et al. (2020) optimize a function that generates a subset of images from a given image dataset, such as both data achieving a similar distribution. GANs have been used for image inpainting, style transfer, and image-to-image translation, among others Jiang et al. (2020); Zhu et al. (2017); Isola et al. (2017); Brock et al. (2018); Choi et al. (2018).

Given a dataset $\boldsymbol{x}^{(i)}{}_{i=1}^{K}$ of K images, it is assumed to follow a distribution $p\text{data}(\boldsymbol{x})$, which is unknown. Subsequently, a generative model $\mathscr{G}$ (DNN) learns the distribution p_g from the data $\boldsymbol{x}$. In this generative model, a prior is assumed on input noise variables $p_{\boldsymbol{z}}(\boldsymbol{z})$ with a known distribution, which is then mapped to the desired generated images. Additionally, a discriminative model $\mathscr{D}$ (another DNN) is defined, which receives the generated samples or dataset samples. GANs consist of two models that engage in an adversarial training process: a generative model $\mathscr{G}$ aims to generate realistic samples to deceive a discriminator model $\mathscr{D}$, which in turn attempts to distinguish between genuine samples from a dataset and the synthetic samples created by the generator. This is called

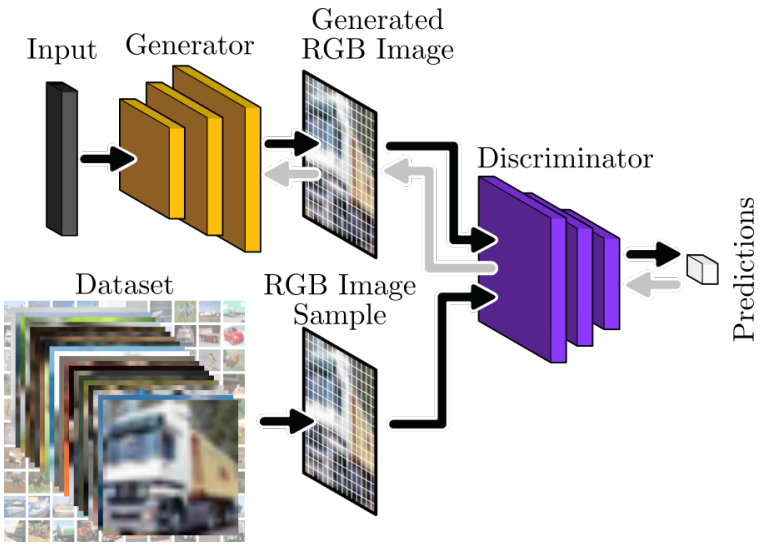

Figure 8

Adversarial training framework. The generative model generates fake images from random noise. The discriminator model tries to identify when the input image is real (from the given dataset) or not (from the generator model).

a min-max game and the optimization problem can be represented as

$$\arg\min_{\mathscr{G}} \arg\max_{\mathscr{D}} \mathbb{E}_{\boldsymbol{x}\sim p_{data}(\boldsymbol{x})}[\log(\mathscr{D}(\boldsymbol{x}))] + \mathbb{E}_{\boldsymbol{z}\sim p_{\boldsymbol{z}}(\boldsymbol{z})}[\log(1-\mathscr{D}(\mathscr{G}(\boldsymbol{z})))]. \tag{4}$$

From the first adversarial training framework Goodfellow et al. (2020), shown in Fig. 8, new types of GANs were introduced considering the dataset information, such as conditional GANs Mirza and Osindero (2014); Isola et al. (2017), where the input of the generative model is conditioned by the label belonging to the designed class for the generated image. Wassertein GAN (W-GAN) Arjovsky et al. (2017) which involves the wasserstein distance metric. W-GAN tries to mitigate the collapse mode present in GANs, where the optimization of the models can converge to a local

optimum, generating the same images independent from the input data. Cycle GAN Zhu et al. (2017) was introduced for style transfer with unpaired datasets, where the objective is to optimize a pair of generative models with their respective discriminator models. The first pair generates the desired image with the transferred style, and the second pair performs the inverse mapping to ensure that the image generation of the first model approximates the initial image. This section explores image generation by GANs for RGB and spectral images.

2.3.1. RGB image Generation. Image generation techniques rely on factors like input data, adversarial model architecture, loss functions, and prior information. Traditional GANs use noise input following a uniform or normal distribution Montgomery and Runger (2010). Some methods incorporate labels as additional input Mirza and Osindero (2014); Isola et al. (2017), while others replace noise with input images for tasks like image-to-image translation Isola et al. (2017); Choi et al. (2018).

Initially, GANs employed fully connected networks for their adversarial models. Subsequently, convolutional neural networks were introduced, often with U-shaped or residual architectures, for both generative and discriminative models Radford et al. (2015); Nemirovsky et al. (2020). More recently, in pursuit of higher-quality image generation, attention mechanisms and self-attention from transformer networks have been integrated into GANs Brock et al. (2018); Park and Kim (2022). Furthermore, advancements have led to the development of models incorporating multiple pairs of generative and discriminative networks, enabling multi-domain image-to-image translation Zhu et al. (2017); Choi et al. (2018).

Loss or regularization functions are introduced in GANs as support or guide elements to an

optimal or desired convergence. To name a few, Arjovsky et al. (2017) introduces the Wasserstein distance to stabilize adversarial training. Cheon et al. (2018) employs a perceptual loss to improve perceptual performance with a slight increase in distortion. Qi (2020) apply a regularization technique based on the Lipschitz regularity condition. Tseng et al. (2021) introduce a regularization term for adversarial training with limited data. Bullwinkel et al. (2022) solves differential equations to learn a loss function for optimizing a DNN.

2.3.2. Spectral Image Generation. Spectral imaging has been studied via GAN architectures for classification tasks as a data augmentation technique. Spectral images or signature samples are generated to improve classification accuracy. Zhan et al. (2017) explore the generation of spectral signature samples based on 1D-GAN. Subsequently, Zhu et al. (2018) employs a GAN to propose two antagonistic networks: a spectral classifier (1D-GAN) and a spectral-spatial classifier (robust 3D-GAN). Zhong et al. (2019) uses a conditional random field framework to refine the classification maps. Feng et al. (2020) propose collaborative learning with an attention mechanism, where generator and discriminator models with attention modules share information. Due to the class imbalance of spectral images, GANs associate minority class samples with false labels. Thus, He et al. (2022) propose a semi-supervised GAN with transformers. Liu et al. (2020b) propose a conditional cascade GAN for the generation of hyperspectral samples, where the first stage model generates spatial information and the second stage generates spatial-spectral information from the generated spatial region. Notice that these approaches involve spectral patches instead of generating complete spectral images.

To obtain complete spectral images, Hauser et al. (2021) proposed an adversarial training

for generating a synthetic dataset of spectral images using RGB images based on the Wasserstein distance metric. Zhu et al. (2021) proposed an unsupervised approach for spectral image reconstruction from RGB images employing priors such as semantic information. Despite the spectral image generation capabilities of these models, they are just mapping RGB images to spectral images in an unsupervised manner. The most similar approach to this research is Martinez et al. (2023), where a generator model can generate spectral images from random input noise based on adversarial training and latent representation of spectral images. Even though, generated spectral images show a low spatial coherence, which is not desirable in GAN models. Unlike currently proposed methods, this research generates spectral images from random noise guided by an RGB image dataset.

3. RGB-guided Spectral Image Generative Adversarial Network

This chapter introduces the proposed method, focusing on spectral implicit learning for knowledge transfer through weight adjustment across different domain image datasets. The concept of spectral implicit learning will be explained, providing insights into how knowledge is transferred. Subsequently, the proposed method will be detailed, highlighting spectral image generation guided by an RGB image dataset and spectral implicit learning. Finally, a spectral band post-processing step will be executed to enhance the robustness of the generated spectral images.

3.1. Spectral Implicit Learning

Since DNNs are powerful information learners, they can be pretrained to be employed in different computational tasks through well-known techniques such as transfer learning or domain adaptation . In this research, spectral image reconstruction is considered to learn a model that can map spectral information in an end-to-end manner, as shown in Fig. 9. Specifically, given an DNN $\mathscr{M}(\cdot)$, the objective is to solve the following optimization problem

$$\mathscr{M}^* = \arg\min_{\mathscr{M}} \mathbb{E}_{\boldsymbol{x}\sim p_{spec}(\boldsymbol{x})}[||\boldsymbol{x} - \mathscr{M}(\boldsymbol{x})||_2^2], \tag{5}$$

where p_{spec} is the spectral image dataset distribution and $\hat{\mathscr{M}}$ is the optimized model. The pretrained model $\hat{\mathscr{M}}$ can be used to implicitly transfer the spectral information to the computational that will be adapted in the next subsection.

The equation (5) is theorically related to the fixed-point method, an iterative algorithm for

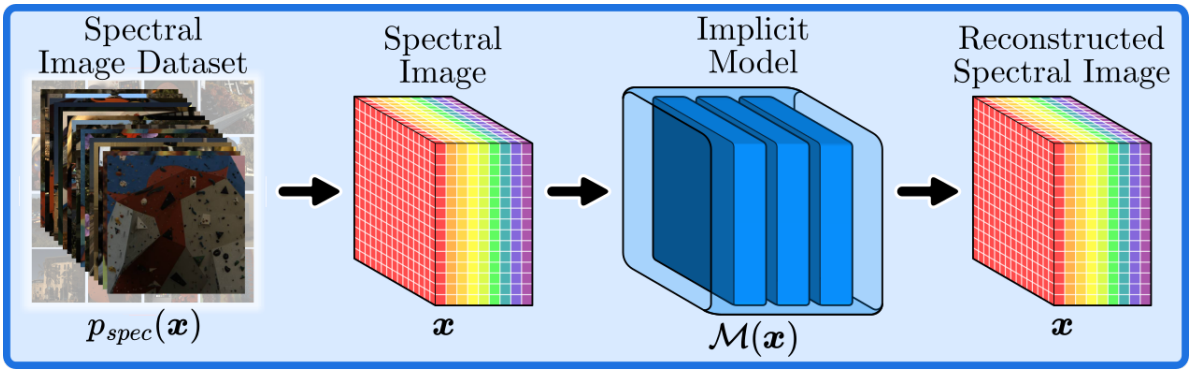

Figure 9

Spectral implicit learning. A DNN $\mathscr{M}$ called implicit model is trained in an end-to-end manner using a spectral image dataset to recover itself.

finding roots Mathews (2004). This algorithm identifies the root of the form $x = f(x)$, where $f(\cdot)$ is the target function. Then, the iterative process is described as $x_{k+1} = f(x_k)$ for $k = 0, 1, \cdots$. Now, if we reinterpret $f(\cdot)$ as a function with optimizable parameters $f_\theta(\cdot)$, and consider a dataset $\{x^{(i)}\}_{k=1}^{K}$ with K paired samples, we can modify the behavior of the function to find parameters that approximate the roots of the function itself by solving the optimization problem

$$\theta^* = \arg\min_{\theta} \sum_{k=1}^{K} ||x^{(i)} - f_\theta(x^{(i)})||_2^2. \tag{6}$$

This implies that the proposed implicit function in (8) aims to find the fixed points of the form $\boldsymbol{x} = \mathscr{M}^*(\boldsymbol{x})$, where the most relevant information about any $\boldsymbol{x}$ is captured through its optimized parameters.

3.2. RGB-guided Spectral Image Generation with Implicit Learning

The proposed method is guided by an RGB image dataset to generate spectral images. Specifically, given a dataset of $\{\boldsymbol{y}^{(i)}\}_{i=1}^{T}$ RGB images with T samples and a spectral response function $\boldsymbol{R}$, the proposed method can learn to generate spectral images through the following adversarial loss

$$\mathscr{L}_{gan}(\mathscr{G},\mathscr{D}) = \mathbb{E}_{\boldsymbol{y}\sim p_{rgb}(\boldsymbol{y})}[log\mathscr{D}(\boldsymbol{y})] + \mathbb{E}_{\boldsymbol{z}\sim p_{\boldsymbol{z}}(\boldsymbol{z})}[log(1-\mathscr{D}(\boldsymbol{R}\mathscr{G}(\boldsymbol{z})))], \tag{7}$$

where $\boldsymbol{z}\in\mathbb{R}^{D}$ is a random input noise with D values that belongs to the distribution $p_{\boldsymbol{z}} = \mathscr{N}(\boldsymbol{0},\boldsymbol{I})$ and p_{rgb} is the distribution of the RGB image dataset. In (7), spectral images are generated by the generator model as $\hat{\boldsymbol{x}} = \mathscr{G}(\boldsymbol{z})$. Posteriorly, the generated spectral image is projected to the RGB space using the spectral response function as $\hat{\boldsymbol{y}} = \boldsymbol{R}\hat{\boldsymbol{x}}$. Finally, the discriminator model determines if the received RGB image is real or fake (belongs to p_{rgb} or p_g, respectively), as shown in Fig. 10. Since there is not any prior information that guided the equation (7) to have realistic spectral information, the spectral response function $\boldsymbol{R}$ can guide the function to have any spectral information pattern, which is undesirable. Therefore, it is necessary to introduce the spectral information through some strategy. In this research, an implicit learning strategy is developed to achieve this purpose. Specifically, an implicit regularization loss is proposed to inject the missing spectral information indirectly as

$$\mathscr{L}_{implicit}(\mathscr{G},\mathscr{M}^{*}) = \mathbb{E}_{\boldsymbol{z}\sim p_{\boldsymbol{z}}(\boldsymbol{z})}[||\mathscr{G}(\boldsymbol{z}) - \mathscr{M}^{*}(\boldsymbol{R}\mathscr{G}(\boldsymbol{z}))||_2^2]. \tag{8}$$

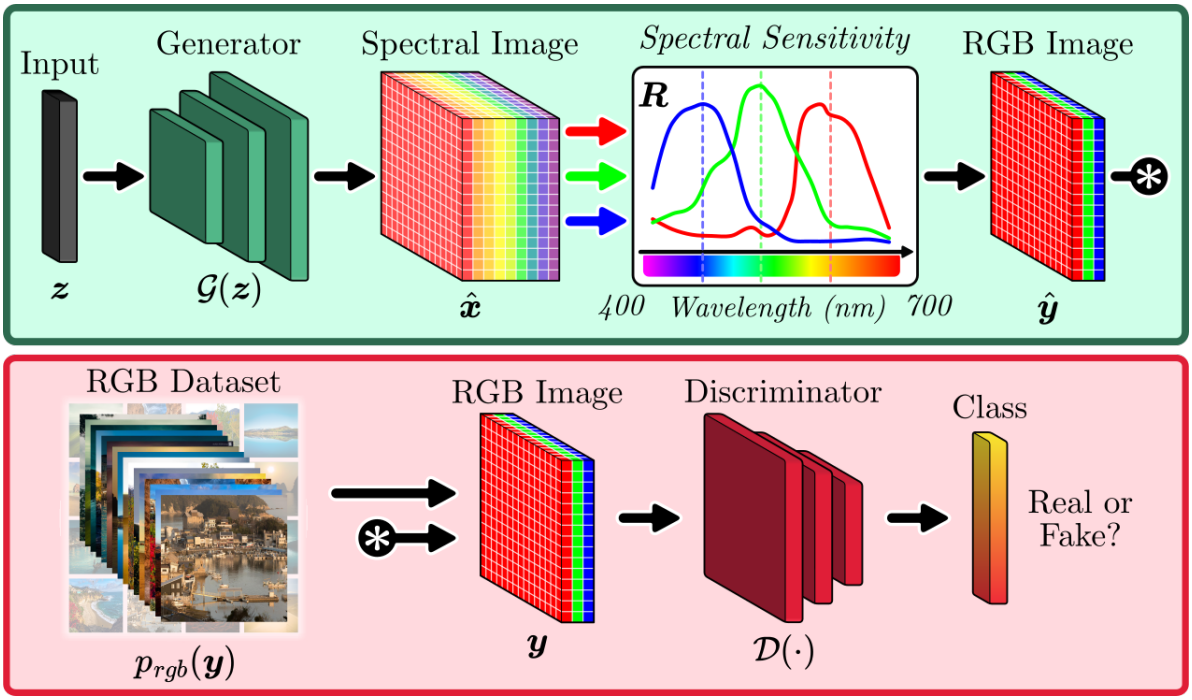

Figure 10

Adversarial training framework. The generative model generates fake spectral images from random noise. The fake spectral images are mapped through a spectral response function ***R*** *to obtain RGB images. The discriminator model tries to identify real images (from the given RGB dataset) and fake images (from the generator model).*

The implicit loss $\mathscr{L}_{implicit}$ guides the generated spectral images to have consistent spectral information employing the pretrained model $\mathscr{M}^*$ with spectral images. Finally, merging the equations (7) and (8), the optimization problem to solve becomes

$$\{\mathscr{G}^*, \mathscr{D}^*\} = \arg\min_{\mathscr{G}} \arg\max_{\mathscr{D}} \mathscr{L}_{gan}(\mathscr{G}, \mathscr{D}) + \lambda \mathscr{L}_{implicit}(\mathscr{G}, \mathscr{M}^*), \tag{9}$$

where λ is a regularization parameter to control the weight relevance of $\mathscr{L}_{implicit}$ in the adversarial training.

3.3. Spectral Band Normalization Post-Processing Step

It is important to emphasize that the normalization of certain spectral signatures between real and generated spectral images is not proportional. This occurrence arises from the unsupervised nature of the proposed implicit regularizer, where both observations and ground truth are unknown. This unsupervised approach impacts the efficacy of validation data for various computational spectral imaging tasks, particularly when dealing with different distributions of spectral normalization.

Therefore, a final post-processing step is applied over generated spectral images considering the mean and standard deviation from the real spectral images. Specifically, if we consider a vectorized spectral image $\boldsymbol{x} \in \mathbb{R}^n$, where $n = MNL$, being (M,N) the spatial resolution and L the number of bands. Then, we can represent each spectral band as $\boldsymbol{x}^l_{(l-1)MN+1:lMN} \in \mathbb{R}^{MN}$, where $l = 1,\cdots,L$. Therefore, considering a generated spectral image $\boldsymbol{x}_g$, we compute their mean and standard deviation for each spectral band as $\boldsymbol{\mu}^l_g = \mathbb{E}[\boldsymbol{x}^l_g] \in \mathbb{R}^L$ and $\boldsymbol{\sigma}^l_g = \sqrt{\mathbb{V}[\boldsymbol{x}^l_g]} \in \mathbb{R}^L$, respectively. The post-processing step is performed as

$$\boldsymbol{x}^l_s = \frac{\boldsymbol{x}^l_g - \boldsymbol{\mu}^l_g}{\boldsymbol{\sigma}^l_g}, \tag{10}$$

$$\boldsymbol{x}^l_a = \boldsymbol{x}^l_s \odot \boldsymbol{\sigma}^l_r + \boldsymbol{\mu}^l_r, \tag{11}$$

where $\boldsymbol{x}_s$ is the standardized spectral image, $\boldsymbol{x}_a$ is the adjusted spectral image, and $\odot$ is the element-wise product. Also, $\boldsymbol{\mu}^l_r = \mathbb{E}[\boldsymbol{x}^l_r] \in \mathbb{R}^L$ and $\boldsymbol{\sigma}^l_r = \sqrt{\mathbb{V}[\boldsymbol{x}^l_r]} \in \mathbb{R}^L$ are the mean and standard deviation

of spectral bands from all the real known spectral images, respectively.

4. Simulations & Results

4.1. Datasets

The selected RGB dataset for guiding the proposed method is landscape Afifi et al. (2021), which contains 4,319 landscape images with different high resolutions. The landscape dataset is divided into 7 different types of landscapes from mountains, deserts, and seas to beaches, islands, and Japan places. The selected spectral dataset is the ARAD1K dataset Arad et al. (2022), which contains 900 samples for training and 50 samples for validation with a spatial resolution of 482 $\times$ 512 and 31 spectral bands. To reduce the computational cost, the spatial resolution of both datasets is reduced to 256 $\times$ 256. For the spectral response function $\boldsymbol{R}$, the provided by Arad et al. (2022) is employed. A visualization of the datasets and the spectral response function are shown in Fig. 11.

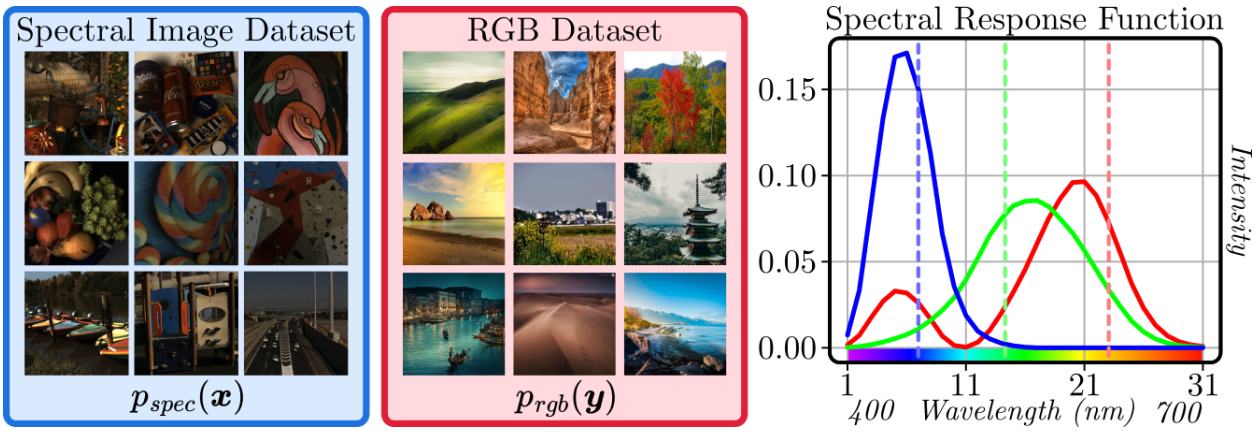

Figure 11

Datasets and spectral response function. From left to right: RGB-mapped Arad1k dataset, landscape dataset, and the spectral response function. As can be seen, the RGB-mapped Arad1k and landscape present different image domains in style and illumination.

Through the development of this research work, it was observed that for a data augmentation strategy, the use of additional training spectral image samples does not improve the CSI method. This can be attributed to the sufficient spatio-spectral diversity present in the original training images. Therefore, five training spectral images were selected from the Arad1k dataset for validating the spatio-spectral quality of the generated spectral images by the proposed method. Justification and results of this strategy are presented in subsection 4.3.3.

4.2. Evaluation Metrics

In this research, there are two relevant groups of metrics to perform a suitable evaluation of the proposed method. The first group are metrics for image generation and the second group are metrics for spectral image reconstruction.

4.2.1. Image Generation Metrics. Unlike other deep learning techniques where a pair of samples with their respective ground truth are known to perform an evaluation using some metric, it is not possible to perform a direct evaluation over the generated samples. Therefore, image generation metrics emerged as a solution to this gap. Salimans et al. (2016) proposes the inception score (IS) metric to compute the statistical difference between the conditional label distribution $p(c|\boldsymbol{x})$ and the marginal probability distribution $p(c)$ as

$$IS = \exp\left(\mathbb{E}_{\boldsymbol{z}\sim p_{\boldsymbol{z}}(\boldsymbol{z})} D_{\mathrm{KL}}(p(c|\boldsymbol{x}=\mathscr{G}(\boldsymbol{z}))||p(c))\right), \tag{12}$$

where $D_{\mathrm{KL}}(\cdot||\cdot)$ is the Kullback-Leibler divergence that enables to compare to statistical distribution with different estimators. Posteriorly, Heusel et al. (2017) proposes de Fréchet Incep-

tion Distance (FID) $d(\cdot,\cdot)$ that compares the first two statistical moments between generated and real RGB images as

$$FID = d^2((\boldsymbol{m},\boldsymbol{C}),(\boldsymbol{m}_g,\boldsymbol{C}_g)) = ||\boldsymbol{m}-\boldsymbol{m}_g||_2^2 + \mathrm{Tr}(\boldsymbol{C}+\boldsymbol{C}_g - 2(\boldsymbol{C}\boldsymbol{C}_g)^{1/2}), \tag{13}$$

where $(\boldsymbol{m},\boldsymbol{C})$ are the mean and covariance matrix obtained from $p_{data}(\cdot)$, and $(\boldsymbol{m}_g,\boldsymbol{C}_g)$ are the mean and covariance matrix obtained from $p_g(\cdot)$.

4.2.2. Spectral Image Reconstruction Metrics. Given a ground truth spectral images $\boldsymbol{x}$ and its reconstruction $\hat{\boldsymbol{x}}$ through some computational algorithm, reconstruction metrics enable to quantify the reconstruction of the spectral images. The employed metrics are the peak signal-to-noise ratio, represented as

$$PSNR = 10 \cdot \log_{10}\left(\frac{\max(\boldsymbol{x})^2}{\sum_{i=1}^{N}(\hat{\boldsymbol{x}}_i - \boldsymbol{x}_i)^2}\right), \tag{14}$$

where $\max_i$ is the maximum value of $\boldsymbol{x}$. The structural similarity index metric allows to predict the quality of a reconstructed image as

$$SSIM = \frac{(2\mu\mu_r + c_1)(2\sigma_m + c_2)}{(\mu^2 + \mu_r^2 + c_1)(\sigma^2 + \sigma_r^2 + c_2)}, \tag{15}$$

where (μ,μ_r) are the mean of a spectral image $\boldsymbol{x}$ and its reconstruction $\hat{\boldsymbol{x}}$, respectively. In the same manner (σ,σ_r) are the variance of a spectral image $\boldsymbol{x}$ and its reconstruction $\hat{\boldsymbol{x}}$, respectively. σ_m is the covariance between $\boldsymbol{x}$ and $\hat{\boldsymbol{x}}$. Finally, c_1 and c_2 are stabilizers to avoid wrong division. The

spectral angle mapper (SAM) is used to quantify the angular different between spectral signatures of ground truth image $\boldsymbol{x}$ against its reconstruction $\hat{\boldsymbol{x}}$. This metric is represented as

$$SAM = \cos^{-1}\left(\frac{\hat{\boldsymbol{x}}^T \boldsymbol{x}}{\sqrt{\sum_{i=1}^{N} \hat{\boldsymbol{x}}_i^2}\sqrt{\sum_{i=1}^{N} \boldsymbol{x}_i^2}}\right) \tag{16}$$

4.3. Experiments & Results

In this section, the performed experiments for spectral implicit learning, RGB-guided spectral image generation, and CSI tasks will be done with their respective results. All experiments were performed with the Pytorch library using an NVIDIA RTX 3090 video card.

4.3.1. Spectral Implicit Learning. For this step, a Unet model Ronneberger et al. (2015) was implemented, as shown in Fig. 12. The model is divided into an encoder and decoder part and contains 4 depth levels. Each block consists of two 2D convolutional layers with their

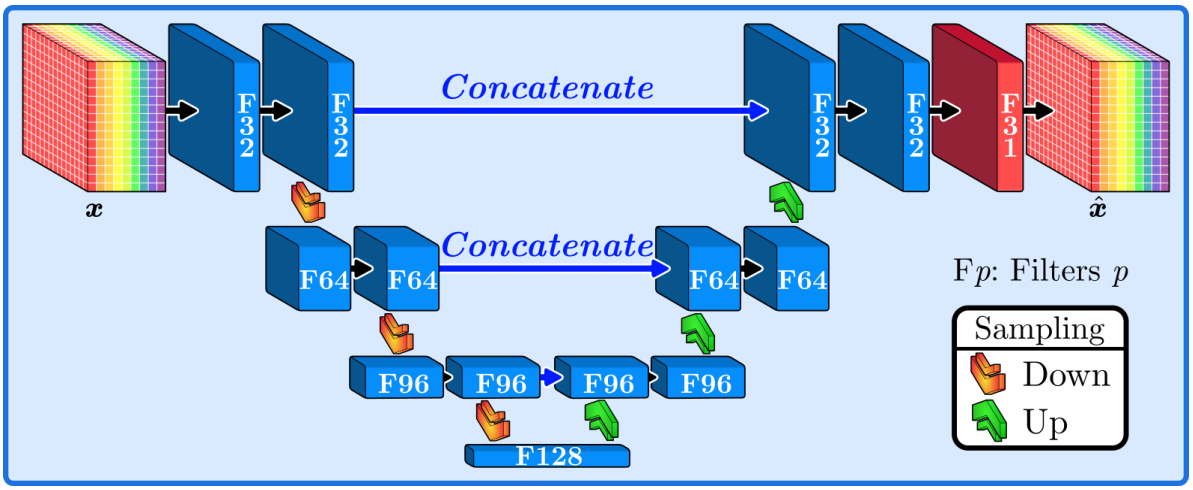

Figure 12

Unet model implemented for the spectral implicit learning.

respective ReLU activations. To reduce the spatial resolution during the encoding part, 2D max pooling operator is applied, and to recover the spatial resolution during the decoding part, 2D upsampling operator is applied. Finally, the last block is a single 2D convolutional layer with its respective ReLU activation. The model setup is a batch size of 2 samples, Adam optimizer with a learning rate of 1×10^{-3} and a reduction on the plateau scheduler with a factor of 0.9 and a patient of 75 epochs during $15,000$ epochs.

Table 1

Performance Metrics of Spectral Image Reconstruction.

Dataset	**PSNR**	**SSIM**	**SAM**
Training	45.68	0.9883	0.0579
Validation	38.63	0.9868	0.0735

The achieved results are shown in table 1, which provides the intuition that the model can learn a wide diversity of spectral information from the 5 training spectral image subset. To go deeper into this intuition, an analysis of the model parameter and spectral reconstruction behavior is performed by doing inference before and after the model training.

From model parameter behavior, a difference can be observed between the untrained and trained Unet model, where a little adjustment can be observed in the model weights, especially when all weights distribution is observed, as shown in Fig. 13. This weight behavior provides the intuition that is only necessary few adjustments in the untrained weights to achieve the weights of the trained model. Then, the model parameter optimization enables to indirect transfer of the necessary spectral information as a guide. This can be corroborated by reconstructing a spectral

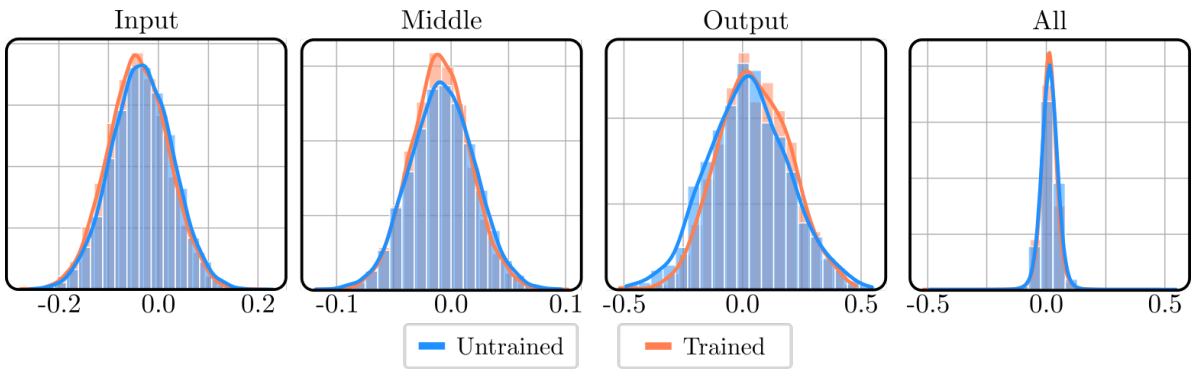

Figure 13

Unet parameters behavior for the untrained and trained model. From left to right: model parameters from input, middle, output, and all the model.

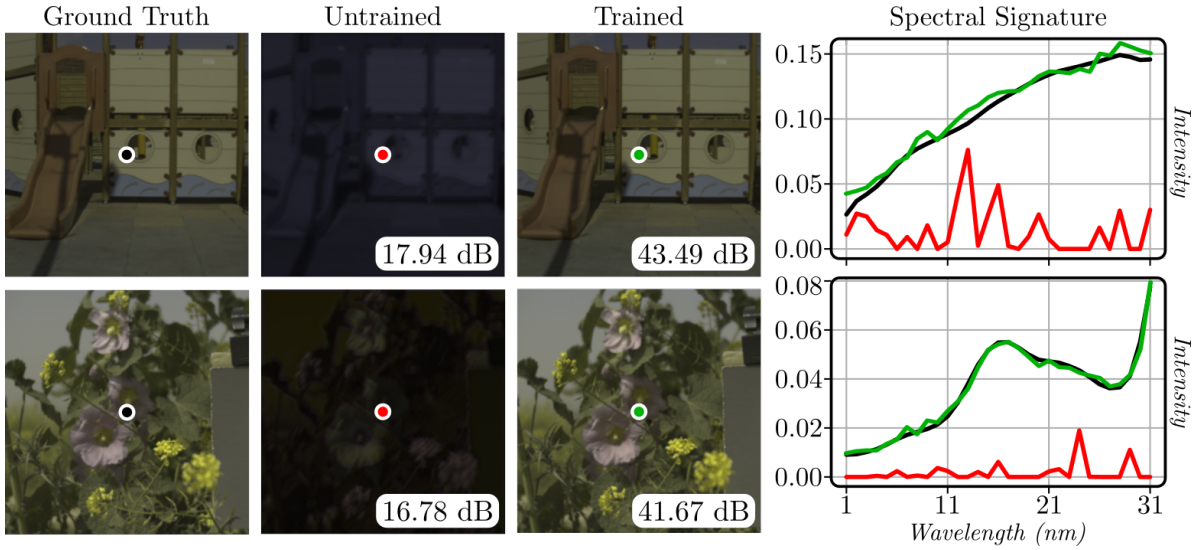

Figure 14

Spectral reconstruction behavior for the untrained and trained model. From left to right: ground truth spectral images, reconstructed spectral images with the untrained and trained Unet model. Finally, the spectral signature from the middle spatial coordinate is represented with black, red, and green colors for the respective spectral images.

image with the untrained and trained model, as shown in Fig. 14, where a spectral signature was plotted from the middle of each spectral image and it can be observed that the untrained model cannot obtain a suitable spectral signature unlike the trained model.

4.3.2. RGB-guided Spectral Image Generation. The proposed method can leverage any GAN from SOTA that generates images from input random noise can be adapted. Then, the projected GAN Sauer et al. (2021) with the Fast GAN Liu et al. (2020a) was selected for the

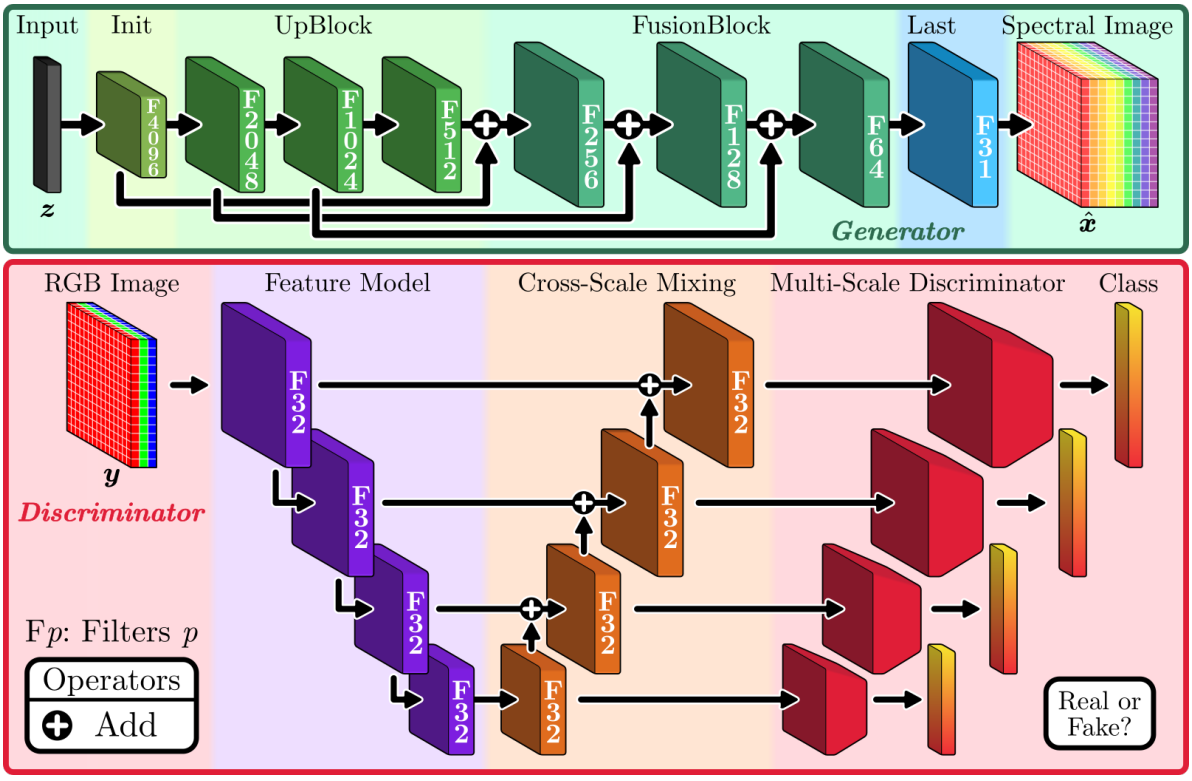

Figure 15

Projected GAN model implemented for the spectral image generation. Top to bottom: the generator model, which generates the spectral image from input random noise. Discriminator model, which discriminates the mapped RGB image from the generated spectral image or real RGB image dataset.

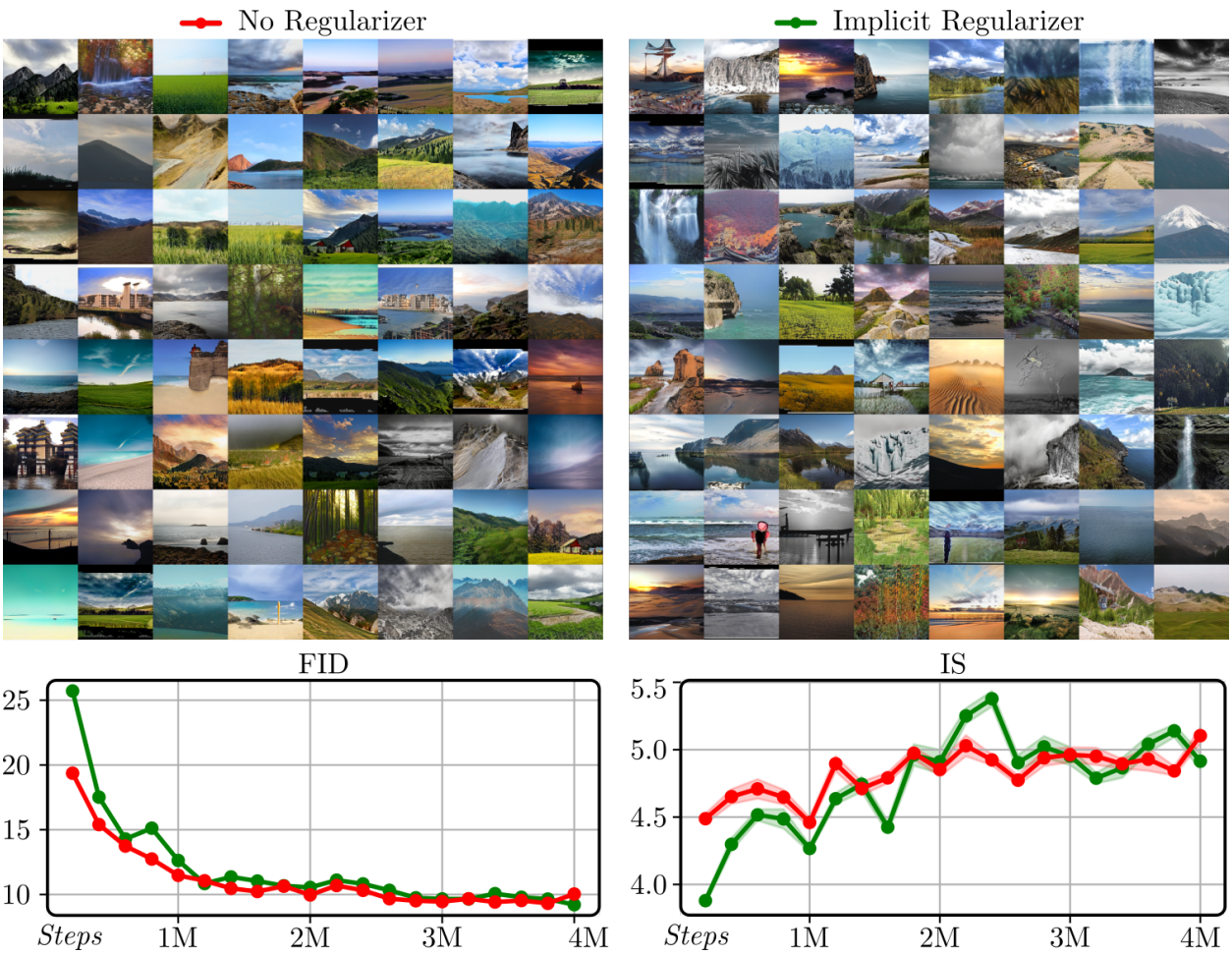

Figure 16

Generated spectral images. From left to right: Generated samples with the proposed method with no regularize and the implicit regularizer. Bottom: FID and IS performance for 4M of gradient steps (M means millions).

proposed method due to its high-quality generation and fast training. The detailed generator and discriminator models are shown in Fig. 15. The generator model is composed of 7 main layers, the first layer is an initialization layer that transforms the random input noise to the first spatial feature map. The following 3 UpBlocks perform upsampling, followed by convolution and GLU

activations. The 3 subsequent Fusion Blocks perform upsampling with feature fusion. Finally, the last layer maps the feature maps to the desired spectral image. Once the spectral image is generated and mapped through the spectral response matrix $\boldsymbol{R}$, the obtained RGB image is passed through the discriminator of the projected GAN. The discriminator model employs a pre-trained classification (feature) model that enables the fast convergence and high-quality generation of the generator model. The feature model is connected in a U-shape with convolutional layers is a Cross-Scale Mixing that enables the feature mapping map to the suitable size for the multi-scale discriminator models to discriminate concerning each feature output, as shown in Fig. 15.

For training, we employ random noise inputs $\boldsymbol{z} \in \mathbb{R}^{64} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I})$, a batch size of 16 samples and a learning rate of 2.5×10^{-3} for the generative model and 2×10^{-3} for the discriminator model, as was implemented in Sauer et al. (2021). Finally, we perform $4M$ of gradient steps per batch where an optimal generation accuracy is achieved (M means millions). The implicit regularization parameter λ from (9) was set to 1×10^3 during the training process.

The performed experiments for spectral image generation are shown in Fig. 16. The generated samples show that the proposed spectral image generation works properly achieving high-quality results. Specifically, observing the FID and IS metrics, both experiments achieve a similar performance. To observe the quality of spectral information, visualization of two real spectral images and 4 generated spectral images by the proposed method are shown in Fig. 17. Three spectral signatures are visualized according to the colors (blue, green, and brown). From generated samples with the regularizers, a high similitude can be observed for each sample, especially for the generated samples with the proposed regularizers.

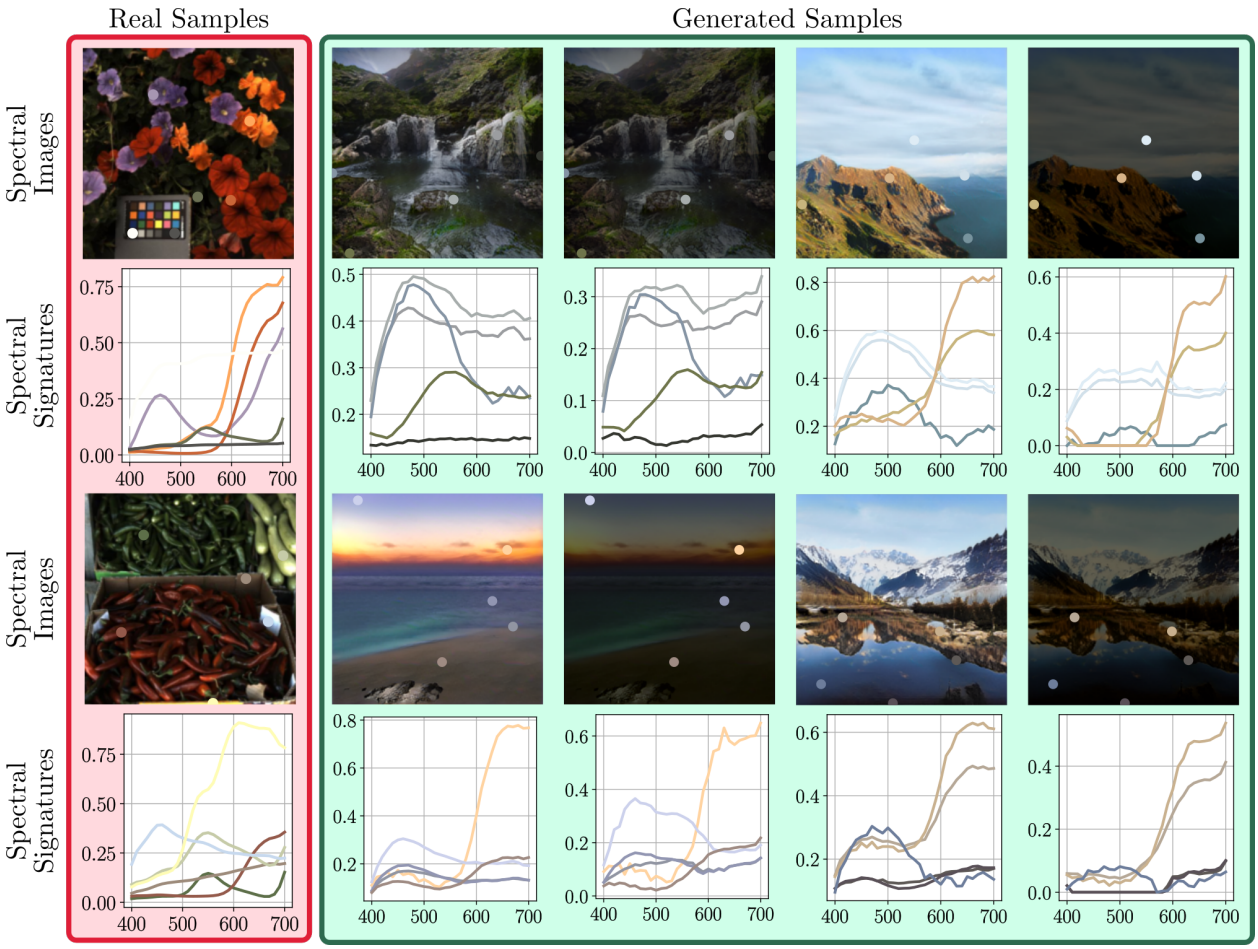

Figure 17

Generated spectral images and their spectral band normalization process. From left to right: Two real spectral images and four generated spectral images with some selected spectral signatures.

4.3.3. Data Augmentation on Compressive Spectral Imaging. To validate the generated spectral images, a data augmentation strategy is applied for CSI, as shown in Fig. 18. For this purpose, the following models were implemented: Unet Ronneberger et al. (2015) and the Deep Spatio-Spectral Prior (DSSP) Huang et al. (2021). DSSP is an unrolled network described in 2.1.2 through the equation (3). The acquisition system is the coded aperture snapshot spectral

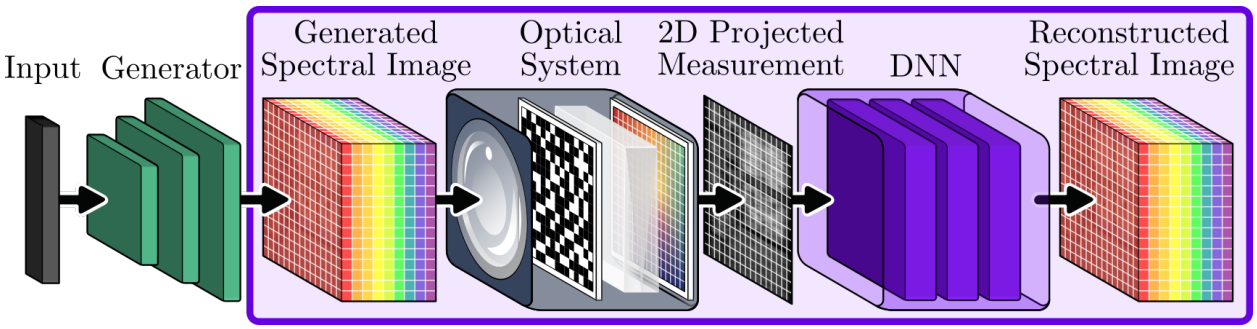

Figure 18

Data augmentation for CSI. A fixed number of spectral images are generated for training a DNN that reconstructs spectral images from measurements.

imaging system (CASSI) Arguello et al. (2023); Carlsson et al. (2024), which performs a spatio-spectral codification of the light providing 2D projected measurements. For the experiments, the training setup consists of 5 real spectral images for baseline results and 900 generated spectral images for the proposed data augmentation strategy. The models are trained exclusively with generated samples to validate their ability to accurately reconstruct spectral images based solely on these generated samples. For comparison purposes, the LD-GAN model is employed, as it is the only model capable of generating a diverse variety of spectral images from random input noise, aligning with the objectives of this research work. Then, 900 spectral images were also generated for training. The hyperparameters for the baseline consist of a batch size of 2 samples and 4 samples for the proposed method with a learning rate of 1×10^{-3} until achieving convergence.

The qualitative results are depicted in Fig. 19, demonstrating that the proposed method excels not only in reconstructing the validation spectral image dataset but also in effectively enhancing the reference results. Notably, when the implicit regularizer is applied with or without the

Model	Method	Enhancements		PSNR (dB)	SSIM (%)	SAM (°)
		Implicit	Statistical			
DSSP	Baseline	-	-	28.46	82.76	0.1563
	Proposed	✓	✓	30.58	86.51	0.1587
		✓	✗	29.78	84.14	0.1764
		✗	✗	21.25	39.20	0.4470
	LD-GAN	-	-	17.17	5.49	0.7789
Unet	Baseline	-	-	23.46	71.35	0.2489
	Proposed	✓	✓	29.59	86.07	0.1945
		✓	✗	27.74	77.69	0.2326
		✗	✗	18.54	49.96	0.4122
	LD-GAN	-	-	14.97	13.51	0.5789

Table 2

Spectral image reconstruction performance on CSI task. The top-performing results are shaded in green and blue, respectively.

statistical band normalization step, improvements are observed in all three reconstructed spectral images presented. Unlike LD-GAN, this method fails to reconstruct the spectral images due to the lack of coherent spatial information in the generated spectral images. The colored dots in the reference results are the spectral signatures plotted on the right side of Fig. 19. From the spectral signature, it can be seen that the proposed method (orange and green lines) achieves a spectral signature that is more similar to the ground truth signature (black line) than the other methods. From the quantitative results shown in Table 2, the proposed method improves the reconstruction of the spectral images in both models DSSP and Unet by at least 2 dBs. Similarly to the qualitative results, the LD-GAN method achieved the worst performance due to the lack of spatial information in the generated train spectral images

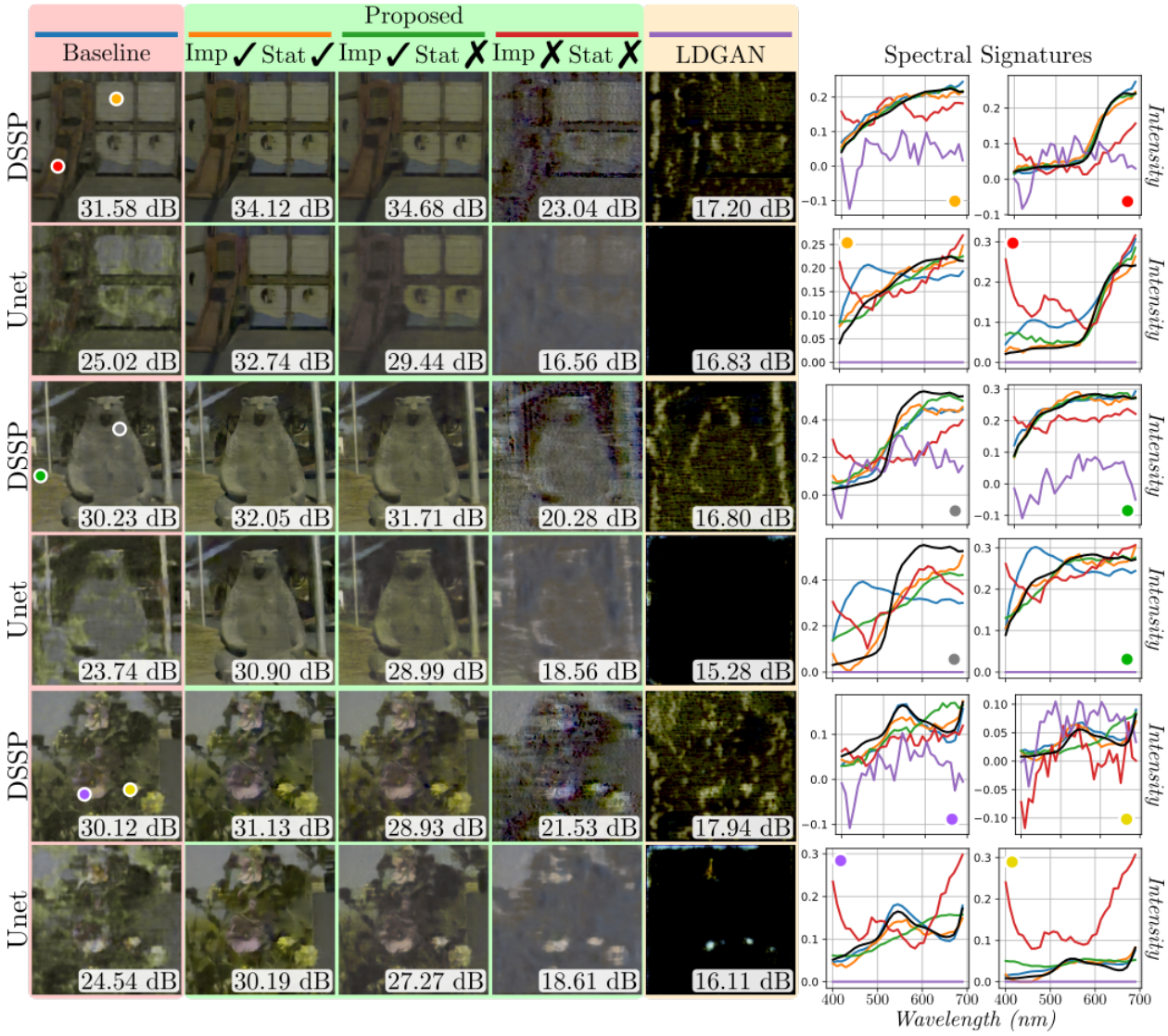

Figure 19

Results on CSI for three samples from the validation set. From left to right: Baseline reconstruction (5 training spectral images) with DSSP and Unet models. Reconstruction using the proposed method using the implicit (Imp) regularize and statistical (Stat) band normalization step. LD-GAN reconstruction results. Finally, spectral signatures from the colored points where each colored line represents the respective method reconstruction (The black line is the ground truth spectral signature).

5. Conclusions

This research work proposed a method to generate spectral images through a GAN guided by RGB images, where the generator samples spectral images from random noise input that are mapped using a spectral response function to RGB images. The RGB images generated and mapped together with the real RGB images of a data set are analyzed by a discriminator model to determine whether the samples are real or fake. Since the spectral response function cannot guarantee authentic spectral information, an implicit regularizer was proposed to inject spectral information into the samples generated through a model pre-trained with spectral images. Finally, a spectral band normalization post-processing step is performed to improve the quality of the generated spectral images. The Arad1k spectral image dataset and the Landscape RGB dataset were employed to guide and generate the spectral images with the proposed method. Experimentally, five training spectral images were taken from the dataset and the results show that it is possible to reconstruct and improve the quality of the reconstruction through a simple DA strategy. The results obtained show that it is not necessary to capture large real spectral images to achieve high-performance results. Finally, the main objective of this research thesis has been accomplished.

Bibliography

Afifi, M., Brubaker, M. A., and Brown, M. S. (2021). Histogan: Controlling colors of gan-generated and real images via color histograms. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 7941–7950.

Arad, B., Timofte, R., Yahel, R., Morag, N., Bernat, A., Cai, Y., Lin, J., Lin, Z., Wang, H., Zhang, Y., et al. (2022). Ntire 2022 spectral recovery challenge and data set. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 863–881.

Arce, G. R., Brady, D. J., Carin, L., Arguello, H., and Kittle, D. S. (2013). Compressive coded aperture spectral imaging: An introduction. *IEEE Signal Processing Magazine*, 31(1):105–115.

Arguello, H. and Arce, G. R. (2014). Colored coded aperture design by concentration of measure in compressive spectral imaging. *IEEE Transactions on Image Processing*, 23(4):1896–1908.

Arguello, H., Bacca, J., Kariyawasam, H., Vargas, E., Marquez, M., Hettiarachchi, R., Garcia, H., Herath, K., Haputhanthri, U., Ahluwalia, B. S., et al. (2023). Deep optical coding design in computational imaging: a data-driven framework. *IEEE Signal Processing Magazine*, 40(2):75–88.

Arjovsky, M., Chintala, S., and Bottou, L. (2017). Wasserstein generative adversarial networks. In *International conference on machine learning*, pages 214–223. PMLR.

Azimi, S. M., Vig, E., Bahmanyar, R., Körner, M., and Reinartz, P. (2018). Towards multi-class

object detection in unconstrained remote sensing imagery. In *Asian conference on computer vision*, pages 150–165. Springer.

Bacca, J., Fonseca, Y., and Arguello, H. (2021). Compressive spectral image reconstruction using deep prior and low-rank tensor representation. *Applied optics*, 60(14):4197–4207.

Bacca, J., Martinez, E., and Arguello, H. (2023). Computational spectral imaging: a contemporary overview. *JOSA A*, 40(4):C115–C125.

Bartczak, P., Iso-Mustajarvi, M., Vrzakova, H., Bednarik, R., Fraunberg, M., and Elomaa, A.-P. (2018). A portable system for on-site medical spectral imaging: pre-clinical development and early evaluation. In *2018 IEEE 31st International Symposium on Computer-Based Medical Systems (CBMS)*, pages 256–261. IEEE.

Beck, A. and Teboulle, M. (2009). A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM journal on imaging sciences*, 2(1):183–202.

Bioucas-Dias, J. M. and Figueiredo, M. A. (2007). A new twist: Two-step iterative shrinkage/thresholding algorithms for image restoration. *IEEE Transactions on Image processing*, 16(12):2992–3004.

Boyd, S., Parikh, N., Chu, E., Peleato, B., Eckstein, J., et al. (2011). Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends® in Machine learning*, 3(1):1–122.

Brock, A., Donahue, J., and Simonyan, K. (2018). Large scale gan training for high fidelity natural image synthesis. *arXiv preprint arXiv:1809.11096*.

Bullwinkel, B., Randle, D., Protopapas, P., and Sondak, D. (2022). Deqgan: Learning the loss function for pinns with generative adversarial networks. *arXiv preprint arXiv:2209.07081*.

Caelles, S., Maninis, K.-K., Pont-Tuset, J., Leal-Taixé, L., Cremers, D., and Van Gool, L. (2017). One-shot video object segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 221–230.

Cao, X., Yue, T., Lin, X., Lin, S., Yuan, X., Dai, Q., Carin, L., and Brady, D. J. (2016). Computational snapshot multispectral cameras: Toward dynamic capture of the spectral world. *IEEE Signal Processing Magazine*, 33(5):95–108.

Carlsson, M., Martinez, E., Vargas, E., and Arguello, H. (2024). Fast matrix inversion in compressive spectral imaging based on a tensorial representation. *Journal of Electronic Imaging*, 33(1):013034–013034.

Chaudhuri, S. and Kotwal, K. (2013). *Hyperspectral image fusion*. Springer.

Cheon, M., Kim, J.-H., Choi, J.-H., and Lee, J.-S. (2018). Generative adversarial network-based image super-resolution using perceptual content losses. In *Proceedings of the European Conference on Computer Vision (ECCV) Workshops*, pages 0–0.

Choi, Y., Choi, M., Kim, M., Ha, J.-W., Kim, S., and Choo, J. (2018). Stargan: Unified generative

adversarial networks for multi-domain image-to-image translation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 8789–8797.

Clancy, N. T., Jones, G., Maier-Hein, L., Elson, D. S., and Stoyanov, D. (2020). Surgical spectral imaging. *Medical image analysis*, 63:101699.

De Biasio, M., Arnold, T., Leitner, R., McGunnigle, G., and Meester, R. (2010). Uav-based environmental monitoring using multi-spectral imaging. In *Airborne Intelligence, Surveillance, Reconnaissance (ISR) Systems and Applications VII*, volume 7668, pages 331–337. SPIE.

Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., and Fei-Fei, L. (2009). Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pages 248–255. Ieee.

Dong, C., Loy, C. C., He, K., and Tang, X. (2015). Image super-resolution using deep convolutional networks. *IEEE transactions on pattern analysis and machine intelligence*, 38(2):295–307.

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., et al. (2020). An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*.

Feng, J., Feng, X., Chen, J., Cao, X., Zhang, X., Jiao, L., and Yu, T. (2020). Generative adversarial networks based on collaborative learning and attention mechanism for hyperspectral image classification. *Remote Sensing*, 12(7):1149.

Fowler, J. E. (2014). Compressive pushbroom and whiskbroom sensing for hyperspectral remote-sensing imaging. In *2014 IEEE international conference on image processing (ICIP)*, pages 684–688. IEEE.

Gastaldi, F., Koh, C. C., Carli, M., Neri, A., and Mitra, S. K. (2005). Compression of videos captured via bayer patterned color filter arrays. In *2005 13th European Signal Processing Conference*, pages 1–4. IEEE.

Gat, N., Subramanian, S., Barhen, J., and Toomarian, N. (1997). Spectral imaging applications: remote sensing, environmental monitoring, medicine, military operations, factory automation, and manufacturing. In *25th AIPR Workshop: Emerging Applications of Computer Vision*, volume 2962, pages 63–77. SPIE.

Gatys, L. A., Ecker, A. S., and Bethge, M. (2015). A neural algorithm of artistic style. *arXiv preprint arXiv:1508.06576*.

Gehm, M. E., John, R., Brady, D. J., Willett, R. M., and Schulz, T. J. (2007). Single-shot compressive spectral imaging with a dual-disperser architecture. *Optics express*, 15(21):14013–14027.

Gelvez, T. and Arguello, H. (2020). Nonlocal low-rank abundance prior for compressive spectral image fusion. *IEEE Transactions on Geoscience and Remote Sensing*, 59(1):415–425.

Gelvez-Barrera, T., Arguello, H., and Foi, A. (2022). Joint nonlocal, spectral, and similarity low-rank priors for hyperspectral–multispectral image fusion. *IEEE Transactions on Geoscience and Remote Sensing*, 60:1–12.

Genser, N., Seiler, J., and Kaup, A. (2020). Camera array for multi-spectral imaging. *IEEE Transactions on Image Processing*, 29:9234–9249.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., and Bengio, Y. (2020). Generative adversarial networks. *Communications of the ACM*, 63(11):139–144.

Gu, Y., Liu, T., Gao, G., Ren, G., Ma, Y., Chanussot, J., and Jia, X. (2021). Multimodal hyperspectral remote sensing: An overview and perspective. *Science China Information Sciences*, 64(2):1–24.

Hagen, N. A. and Kudenov, M. W. (2013). Review of snapshot spectral imaging technologies. *Optical Engineering*, 52(9):090901.

Hauser, J., Shtendel, G., Zeligman, A., Averbuch, A., and Nathan, M. (2021). Shs-gan: Synthetic enhancement of a natural hyperspectral database. *IEEE Transactions on Computational Imaging*, 7:505–517.

He, K., Gkioxari, G., Dollár, P., and Girshick, R. (2017). Mask r-cnn. In *Proceedings of the IEEE international conference on computer vision*, pages 2961–2969.

He, Z., Xia, K., Ghamisi, P., Hu, Y., Fan, S., and Zu, B. (2022). Hypervitgan: Semisupervised generative adversarial network with transformer for hyperspectral image classification. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 15:6053–6068.

Heusel, M., Ramsauer, H., Unterthiner, T., Nessler, B., and Hochreiter, S. (2017). Gans trained by a two time-scale update rule converge to a local nash equilibrium. *Advances in neural information processing systems*, 30.

Hong, D., Gao, L., Yao, J., Zhang, B., Plaza, A., and Chanussot, J. (2020). Graph convolutional networks for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 59(7):5966–5978.

Huang, T., Dong, W., Yuan, X., Wu, J., and Shi, G. (2021). Deep gaussian scale mixture prior for spectral compressive imaging. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 16216–16225.

Hunt, R. W. G. (1995). *The reproduction of colour*, volume 4. Wiley Online Library.

Isola, P., Zhu, J.-Y., Zhou, T., and Efros, A. A. (2017). Image-to-image translation with conditional adversarial networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1125–1134.

Jiang, Y., Xu, J., Yang, B., Xu, J., and Zhu, J. (2020). Image inpainting based on generative adversarial networks. *IEEE Access*, 8:22884–22892.

Kong, Y. and Fu, Y. (2022). Human action recognition and prediction: A survey. *International Journal of Computer Vision*, 130(5):1366–1401.

Li, H., Ellis, J. G., Zhang, L., and Chang, S.-F. (2018a). Patternnet: Visual pattern mining with

deep neural network. In *Proceedings of the 2018 ACM on international conference on multimedia retrieval*, pages 291–299.

Li, S., Song, W., Fang, L., Chen, Y., Ghamisi, P., and Benediktsson, J. A. (2019). Deep learning for hyperspectral image classification: An overview. *IEEE Transactions on Geoscience and Remote Sensing*, 57(9):6690–6709.

Li, S., Zhang, K., Hao, Q., Duan, P., and Kang, X. (2018b). Hyperspectral anomaly detection with multiscale attribute and edge-preserving filters. *IEEE Geoscience and Remote Sensing Letters*, 15(10):1605–1609.

Li, Y., Wu, C.-Y., Fan, H., Mangalam, K., Xiong, B., Malik, J., and Feichtenhofer, C. (2022). Mvitv2: Improved multiscale vision transformers for classification and detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4804–4814.

Li, Y., Zhang, J., Sun, G., and Wang, S. (2020). Hyperspectral image compressive reconstruction with low-rank tensor constraint. *Journal of Electronic Imaging*, 29(2):023009.

Liang, J., Cao, J., Sun, G., Zhang, K., Van Gool, L., and Timofte, R. (2021). Swinir: Image restoration using swin transformer. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 1833–1844.

Liu, B., Zhu, Y., Song, K., and Elgammal, A. (2020a). Towards faster and stabilized gan training

for high-fidelity few-shot image synthesis. In *International Conference on Learning Representations*.

Liu, X., Qiao, Y., Xiong, Y., Cai, Z., and Liu, P. (2020b). Cascade conditional generative adversarial nets for spatial-spectral hyperspectral sample generation. *Science China Information Sciences*, 63(4):1–16.

Liu, Z., Luo, P., Wang, X., and Tang, X. (2015). Deep learning face attributes in the wild. In *Proceedings of the IEEE international conference on computer vision*, pages 3730–3738.

Luiten, J., Voigtlaender, P., and Leibe, B. (2018). Premvos: Proposal-generation, refinement and merging for video object segmentation. In *Asian Conference on Computer Vision*, pages 565–580. Springer.

Martínez, E., Castro, S., Bacca, J., and Arguello, H. (2020a). Efficient transfer learning for spectral image reconstruction from rgb images. In *2020 IEEE Colombian Conference on Applications of Computational Intelligence (IEEE ColCACI 2020)*, pages 1–6. IEEE.

Martínez, E., Castro, S., Bacca, J., and Arguello, H. (2020b). Transfer learning for spectral image reconstruction from rgb images. In *IEEE Colombian Conference on Applications in Computational Intelligence*, pages 160–173. Springer.

Martinez, E., Jacome, R., Hernandez-Rojas, A., and Arguello, H. (2023). Ld-gan: Low-dimensional generative adversarial network for spectral image generation with variance regu-

larization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 265–275.

Mathews, J. (2004). Numerical methods using matlab.

Mehta, N., Shaik, S., Devireddy, R., and Gartia, M. R. (2018). Single-cell analysis using hyperspectral imaging modalities. *Journal of biomechanical engineering*, 140(2).

Meng, Z., Ma, J., and Yuan, X. (2020). End-to-end low cost compressive spectral imaging with spatial-spectral self-attention. In *European Conference on Computer Vision*, pages 187–204. Springer.

Mirza, M. and Osindero, S. (2014). Conditional generative adversarial nets. *arXiv preprint arXiv:1411.1784*.

Montgomery, D. C. and Runger, G. C. (2010). *Applied statistics and probability for engineers*. John wiley & sons.

Nemirovsky, D., Thiebaut, N., Xu, Y., and Gupta, A. (2020). Countergan: generating realistic counterfactuals with residual generative adversarial nets. *arXiv preprint arXiv:2009.05199*.

Pak, M. and Kim, S. (2017). A review of deep learning in image recognition. In *2017 4th international conference on computer applications and information processing technology (CAIPT)*, pages 1–3. IEEE.

Park, J. and Kim, Y. (2022). Styleformer: Transformer based generative adversarial networks

with style vector. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8983–8992.

Qi, G.-J. (2020). Loss-sensitive generative adversarial networks on lipschitz densities. *International Journal of Computer Vision*, 128(5):1118–1140.

Qi, S., Ning, X., Yang, G., Zhang, L., Long, P., Cai, W., and Li, W. (2021). Review of multi-view 3d object recognition methods based on deep learning. *Displays*, 69:102053.

Radford, A., Metz, L., and Chintala, S. (2015). Unsupervised representation learning with deep convolutional generative adversarial networks. *arXiv preprint arXiv:1511.06434*.

Ronneberger, O., Fischer, P., and Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer.

Salamí, E., Barrado, C., and Pastor, E. (2014). Uav flight experiments applied to the remote sensing of vegetated areas. *Remote Sensing*, 6(11):11051–11081.

Salimans, T., Goodfellow, I., Zaremba, W., Cheung, V., Radford, A., and Chen, X. (2016). Improved techniques for training gans. *Advances in neural information processing systems*, 29.

Sauer, A., Chitta, K., Müller, J., and Geiger, A. (2021). Projected gans converge faster. *Advances in Neural Information Processing Systems*, 34:17480–17492.

Shaw, G. A. and Burke, H. K. (2003). Spectral imaging for remote sensing. *Lincoln laboratory journal*, 14(1):3–28.

Su, J.-W., Chu, H.-K., and Huang, J.-B. (2020). Instance-aware image colorization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7968–7977.

Sun, J., Li, H., Xu, Z., et al. (2016). Deep admm-net for compressive sensing mri. *Advances in neural information processing systems*, 29.

Tseng, H.-Y., Jiang, L., Liu, C., Yang, M.-H., and Yang, W. (2021). Regularizing generative adversarial networks under limited data. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7921–7931.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., and Polosukhin, I. (2017). Attention is all you need. *Advances in neural information processing systems*, 30.

Wang, L., Sun, C., Fu, Y., Kim, M. H., and Huang, H. (2019). Hyperspectral image reconstruction using a deep spatial-spectral prior. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8032–8041.

Wang, Y. W., Reder, N. P., Kang, S., Glaser, A. K., and Liu, J. T. (2017). Multiplexed optical imaging of tumor-directed nanoparticles: a review of imaging systems and approaches. *Nanotheranostics*, 1(4):369.

Wang, Z., Cun, X., Bao, J., Zhou, W., Liu, J., and Li, H. (2022). Uformer: A general u-shaped transformer for image restoration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 17683–17693.

Xiang, J., Dong, Y., and Yang, Y. (2021). Fista-net: Learning a fast iterative shrinkage thresholding network for inverse problems in imaging. *IEEE Transactions on Medical Imaging*, 40(5):1329–1339.

Yuan, X., Brady, D. J., and Katsaggelos, A. K. (2021). Snapshot compressive imaging: Theory, algorithms, and applications. *IEEE Signal Processing Magazine*, 38(2):65–88.

Zhan, Y., Hu, D., Wang, Y., and Yu, X. (2017). Semisupervised hyperspectral image classification based on generative adversarial networks. *IEEE Geoscience and Remote Sensing Letters*, 15(2):212–216.

Zhang, R., Isola, P., and Efros, A. A. (2016). Colorful image colorization. In *European conference on computer vision*, pages 649–666. Springer.

Zhang, Y., Li, K., Li, K., Wang, L., Zhong, B., and Fu, Y. (2018). Image super-resolution using very deep residual channel attention networks. In *Proceedings of the European conference on computer vision (ECCV)*, pages 286–301.

Zhong, Z., Li, J., Clausi, D. A., and Wong, A. (2019). Generative adversarial networks and conditional random fields for hyperspectral image classification. *IEEE transactions on cybernetics*, 50(7):3318–3329.

Zhu, J.-Y., Park, T., Isola, P., and Efros, A. A. (2017). Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Proceedings of the IEEE international conference on computer vision*, pages 2223–2232.

Zhu, L., Chen, Y., Ghamisi, P., and Benediktsson, J. A. (2018). Generative adversarial networks for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 56(9):5046–5063.

Zhu, Z., Liu, H., Hou, J., Zeng, H., and Zhang, Q. (2021). Semantic-embedded unsupervised spectral reconstruction from single rgb images in the wild. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 2279–2288.

Zimmermann, T., Rietdorf, J., and Pepperkok, R. (2003). Spectral imaging and its applications in live cell microscopy. *FEBS letters*, 546(1):87–92.