

Article

Estimating the Spectral Response of Eight-Band MSFA One-Shot Cameras Using Deep Learning

Pierre Gouton ^{1,*}, Kacoutchy Jean Ayikpa ^{1,2}  and Diarra Mamadou ¹

¹ Laboratoire Imagerie et Vision Artificielle (ImVia), Université de Bourgogne, 21000 Dijon, France; kacoutchy-jean.ayikpa@u-bourgogne.fr or kacoutchy.ayikpa@uvci.edu.ci (K.J.A.); patoudiarra@gmail.com (D.M.)

² Unité de Recherche et d'Expertise Numérique (UREN), Université Virtuelle de Côte d'Ivoire, 28 BP 536, Abidjan 28, Côte d'Ivoire

* Correspondence: pgouton@u-bourgogne.fr; Tel.: +33-782194200

Abstract: Eight-band one-shot MSFA (multispectral filter array) cameras are innovative technologies used to capture multispectral images by capturing multiple spectral bands simultaneously. They thus make it possible to collect detailed information on the spectral properties of the observed scenes economically. These cameras are widely used for object detection, material analysis, and agronomy. The evolution of one-shot MSFA cameras from 8 to 32 bands makes obtaining much more detailed spectral data possible, which is crucial for applications requiring delicate and precise analysis of the spectral properties of the observed scenes. Our study aims to develop models based on deep learning to estimate the spectral response of this type of camera and provide images close to the spectral properties of objects. First, we prepare our experiment data by projecting them to reflect the characteristics of our camera. Next, we harness the power of deep super-resolution neural networks, such as very deep super-resolution (VDSR), Laplacian pyramid super-resolution networks (LapSRN), and deeply recursive convolutional networks (DRCN), which we adapt to approximate the spectral response. These models learn the complex relationship between 8-band multispectral data from the camera and 31-band multispectral data from the multi-object database, enabling accurate and efficient conversion. Finally, we evaluate the images' quality using metrics such as loss function, PSNR, and SSIM. The model evaluation revealed that DRCN outperforms others in crucial performance. DRCN achieved the lowest loss with 0.0047 and stood out in image quality metrics, with a PSNR of 25.5059, SSIM of 0.8355, and SAM of 0.13215, indicating better preservation of details and textures. Additionally, DRCN showed the lowest RMSE 0.05849 and MAE 0.0415 values, confirming its ability to minimize reconstruction errors more effectively than VDSR and LapSRN.

Keywords: spectral response; MSFA one-shot; deep learning; multispectral image



Citation: Gouton, P.; Ayikpa, K.J.; Mamadou, D. Estimating the Spectral Response of Eight-Band MSFA One-Shot Cameras Using Deep Learning. *Algorithms* **2024**, *17*, 473. <https://doi.org/10.3390/a17110473>

Academic Editors: Chih-Lung Lin, Bor-Jiunn Hwang, Shaou-Gang Miaou and Chi-Hung Chuang

Received: 29 August 2024

Revised: 1 October 2024

Accepted: 15 October 2024

Published: 22 October 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Acquiring high-spectral-resolution multispectral images from low-resolution sensors represents a significant challenge in various fields, including remote sensing, medical imaging, precision agriculture, and scene analysis [1]. High-resolution multispectral images provide a more detailed and accurate analysis of observed scenes, which is important for environmental monitoring typology and disease-detection applications [2]. The acquisition of multispectral images through systems based on CRI, Hypspx, etc., provides images of 30 to 100 bands, offering unparalleled information [3]. Each spectral band makes it possible to detect specific characteristics of objects that are not visible to the naked eye. As in precision agriculture, these images can help identify areas requiring particular treatments. They can detect tissue abnormalities in healthcare and find applications in many areas. However, its devices are not fast and poorly suited to slow-motion objects. 8-band one-shot MSFA (multispectral filter array) cameras provide a compact, cost-effective solution for spectral

data acquisition. However, they produce images of low spectral resolution. This band limitation makes capturing the details needed for in-depth analyses difficult. Therefore, although these sensors are accessible, the quality of the data they provide is insufficient for many critical applications. Thus, as part of the European CAVIAR project, which aims to improve image-detection technologies in various sectors, notably health, agriculture, etc., by relying on advanced imaging techniques, the use of an 8-band MSFA filter was used for the camera setup to capture a broader spectrum of light [4,5]. However, the transition to a high-spectral-resolution system (31 to 100 bands) requires more sophisticated and expensive sensors, which are slow, limiting their accessibility. This is why estimating the spectral response of 8-band one-shot MSFA cameras raises significant technical challenges. In this context, deep learning methods are a promising solution. This study focuses on applying advanced deep learning techniques to improve the quality and accuracy of the spectral response of 8-band one-shot cameras to approximate hyperspectral systems. By harnessing the power of convolutional neural networks (CNN) and other deep neural-network architectures, it is possible to transform data captured by low-resolution sensors into high-quality multispectral images [6]. This study aims to demonstrate the potential of deep learning techniques to approximate the spectral response of 8-band one-shot cameras, making this technology more accessible and applicable to a wide range of domains. The results obtained open new perspectives for improving sensor systems and image analysis methods, thus contributing to significant advances in research and practical applications.

The specific contributions of our study are as follows:

- We have implemented image datasets that contain 8-band images as input and 31-band images as output. We retrieved the images from the TokyoTech [7] database, which offers 31-band multispectral images. Using a Gaussian projection, we generated 8-band images from this data. These images are then mosaicked and demosaicked to achieve a quality comparable to the camera developed as part of the CAVIAR project. This approach creates a robust dataset for training and evaluating models for estimating the spectral response of 8-band one-shot cameras.
- We have developed several models based on deep-learning algorithms, including very deep super-resolution (VDSR), deeply recursive convolutional networks (DRCN), and Laplacian pyramid super-resolution networks (LapSRN). These models were initially used for super-resolution and then adapted to approximate the spectral response. These models are powerful tools for approximating the spectral response of 8-band one-shot MSFA cameras. By exploiting the advanced architectures of these networks, we aim to significantly improve the quality of the spectral response obtained with these cameras.
- We compare the algorithms used for building the models by evaluating them using specific metrics such as structural similarity index (SSIM) and peak signal-to-noise ratio (PSNR) to measure their performance. Furthermore, a visual comparison between the predicted spectral responses and the ground truths is performed to validate the effectiveness of the proposed methods. This approach allows us to identify the best-performing models and optimize techniques for approximating the spectral response of 8-band one-shot MSFA cameras.

The rest of our manuscript is organized as follows: Section 2 deals with the state of the art regarding the spectral response of multispectral cameras. Section 3 details the materials and methods used in our approach. Section 4 presents the experiments carried out and the results obtained. Section 5 provides a discussion of these results. Section 6 addresses the threats to validity. Finally, Section 7 presents our conclusion.

2. Related Work

Several works have been carried out in the reconstruction of images, and we can cite, among others, Toivonen et al., who developed a practical method for estimating the spectral response of camera sensors and the associated uncertainty using a specific imaging process. This method relies on a restricted set of 15 images, including diffraction

images and color patches of known spectra, to obtain high-resolution spectral response estimates. The approach is versatile and applicable to any camera sensor in the visible domain. It does not impose strict constraints on the possible estimates, thus ensuring high precision of the results, which agree with previous estimates [8]. Darrodi et al. conducted a study to test the performance of different sensor estimation techniques from the literature. They used measured and synthetic data, which they then compared to reference data to evaluate the effectiveness of these techniques on two cameras [9]. Han et al. proposed an estimation method based on a single image and exploiting fluorescence without requiring prior knowledge of the illumination spectrum. Under different light sources, the spectral distributions of the fluorescence emitted by the same material remain constant, up to one scale. Thus, the camera response to fluorescence maintains the same chromaticity. By relying on this chromatic invariance, they estimated the camera's spectral sensitivity under arbitrary illumination, the spectrum of which is unknown. Through numerous experiments, they demonstrated the accuracy of their method under various lighting conditions. Additionally, they showed how to recover daylight spectra from the estimated results and then use the cameras' estimated spectral sensitivities and daylight spectra to solve color correction problems [10]. Zhao et al. studied spectral features using singular value decomposition (SVD) to extract basis functions. They collected data from the literature and direct measurements of the sensitivity of different cameras. Their paper compares the extracted basis functions to identify the optimal set with other mathematical basis functions. These optimized basis functions can then be used to estimate the unknown spectral sensitivity of any camera [11]. Prasad et al. proposed a quick and straightforward method to obtain a reasonable approximation of camera-response functions using only a color chart and unknown, random lighting. This approach is made possible by carefully designing the cost function, which imposes several constraints to make the problem feasible. Among the components of this cost function, the Luther condition serves as a global shape prior, while commercial lighting systems with unknown spectra provide narrow spectral windows for local reconstruction. The quality of the rebuilding obtained is comparable to that of methods using known lighting [12]. Matanga et al. developed an innovative method to approximate the spectral response of color images. They introduced an approach based on neural networks, aiming to improve the two previous methods: those using circular and exponential functions and those based on the inverse of Penrose or Wiener. This new method can also be applied to calibrating most color scanning systems and subwavelength filters [13].

When we talk about a camera's spectral response, we also talk about the quality and reconstruction of the images. Chunwei et al. thus conducted a comparative study on GANs from different angles to address these issues. They began by reviewing recent advances in GANs and then presented popular architectures suitable for image super-resolution, both for small and large samples. They analyzed the motivations, implementations, and differences of optimization methods based on GANs and discriminative learning, including supervised, semi-supervised, and unsupervised approaches. In addition, a comparison of the performance of these GANs was carried out on public datasets, accompanied by a quantitative and qualitative analysis [14]. For their part, Vizilter et al. proposed a new approach to interconnect multispectral images via diffusive morphology techniques to improve aviation vision systems. Unlike traditional morphological methods, this approach does not require image segmenting to describe its shape and offers strong resistance to noise and distortions. They developed procedures for interconnecting two- and three-channel images using diffusive morphological filtrations and an image pyramid. A promising prototype of this improved vision system was tested with multispectral data collected during an aeronautical test [15]. Andriyanov et al. focused on the problem of filtering multispectral images by studying the effectiveness of filtering image sequences. They analyzed variations in error variance as a function of the number of images in the sequence and the inter-image correlation coefficient. To statistically model these images, they used doubly stochastic random fields [16]. Furthermore, Andriyanov et al. explored the restoration of

images under conditions where a portion of the observations is subject to additive noise, or more precisely, the restoration of randomly sampled images with pixels affected by white Gaussian noise. They developed nonlinear filters based on deep doubly stochastic Gaussian models, which were more effective than linear methods and classical algorithms. These filters allow images to be restored using only 50% of the data with a relative error rate of 9% [17]. These studies show the importance of robust and innovative approaches to improve image quality, particularly in multispectral vision and image reconstruction, where precision and resistance to noise are essential.

Most existing methods for approximating spectral responses must be adapted to our specific problem of 8-band MSFA one-shot cameras. Our article presents a new approach based on deep learning to address this challenge.

3. Materials and Methods

The experiments were conducted using Python programming at the ImViA laboratory on a DELL desktop computer. This system has an Intel(R) Core i7-10700 CPU running at 2.90 GHz, with 32 GB of RAM and an NVIDIA Quadro P400 GPU. The models were configured in Python version 3.8.8, utilizing Keras API version 2.4.3 and TensorFlow version 2.3 as the backend, with CUDA 10.1 and cuDNN 7.6 dependencies for GPU acceleration.

3.1. Dataset

Our dataset comes from the multispectral database called TokyoTech. This dataset contains 31-band multispectral images from 420 nm to 720 nm at 10 nm intervals. Images were captured using a monochrome camera with a VariSpec liquid crystal tunable filter (VariSpec VIS). This dataset is composed of 35 different objects, each containing 31 spectral bands [7].

Figure 1 shows an extract of some images from the database.

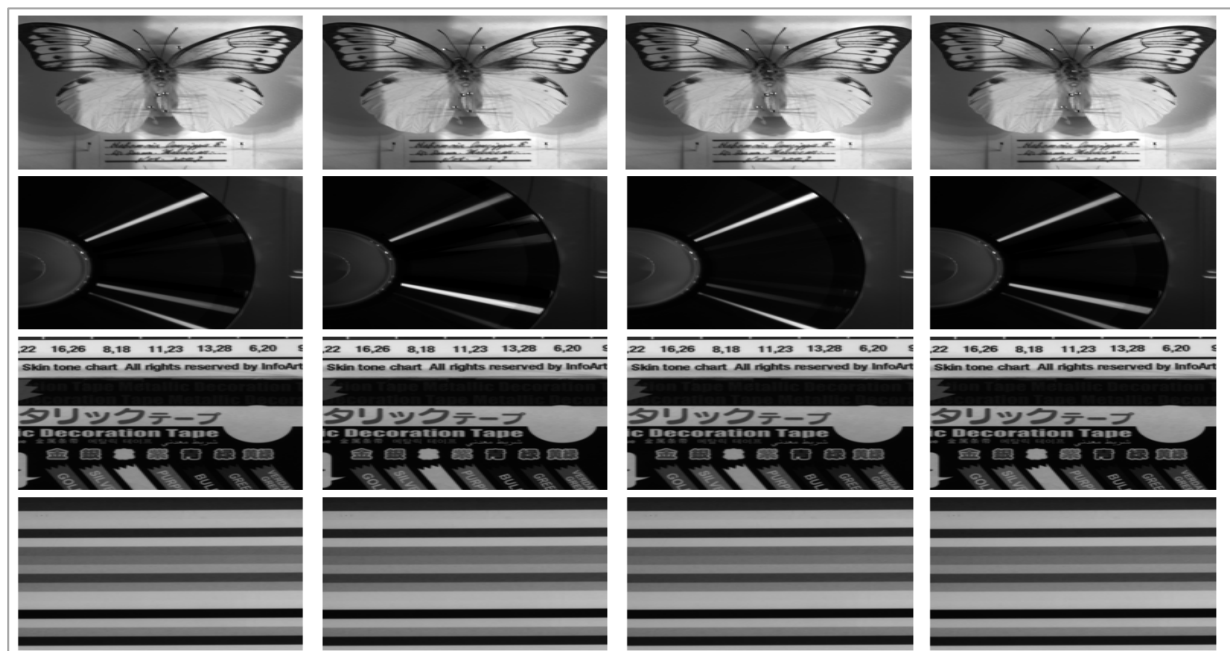


Figure 1. Extracted from the TokyoTech database [7].

We divided each image into 4 segments from this dataset to create a more extensive database for our experiment. This will allow us to obtain a dataset of 140 images containing 31 spectral bands. Figure 2 shows some images from the new dataset.

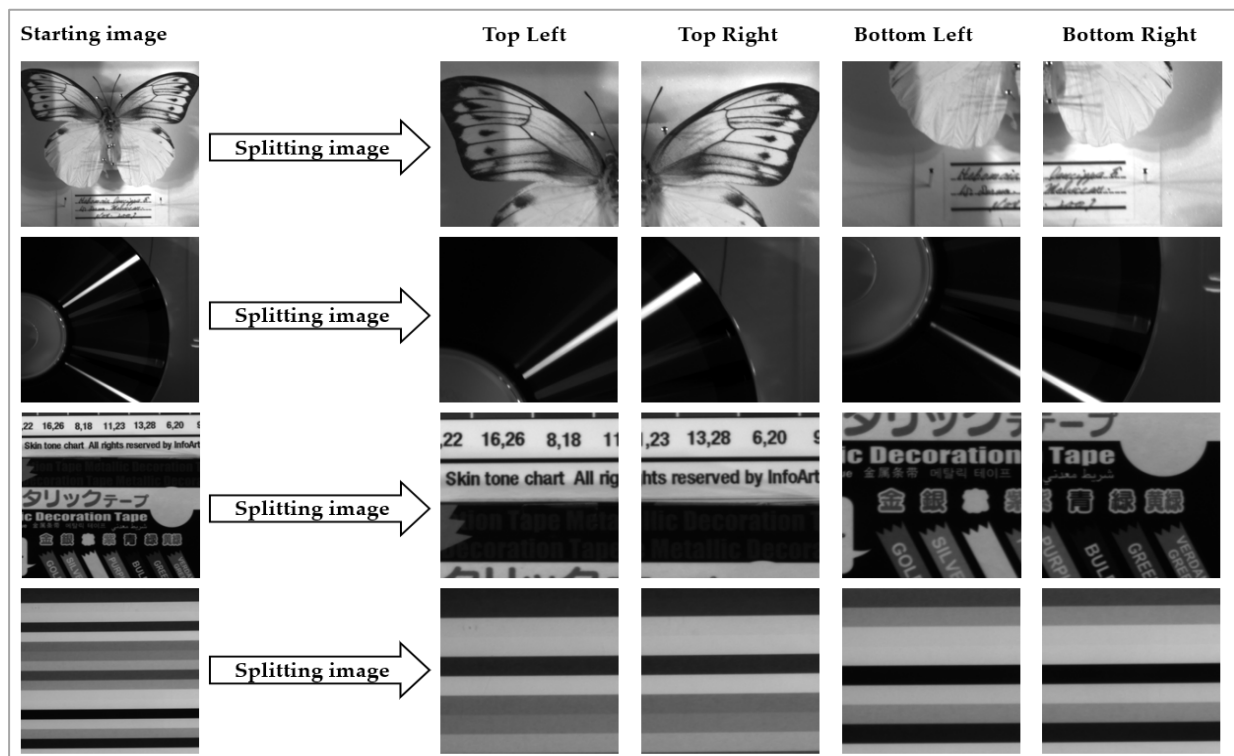


Figure 2. Illustration of split image.

After splitting the data, we used a Gaussian projection, an ideal method to model the MSFA filters of the color shade hybrid sensor and obtain spectral responses from 31 to 8 bands. This projection reduces the dimensionality of the data while retaining essential spectral characteristics, thus facilitating the precise estimation of spectral bands from a reduced number of initial bands. Figure 3 shows the transformation of the 31 bands into 8 bands.

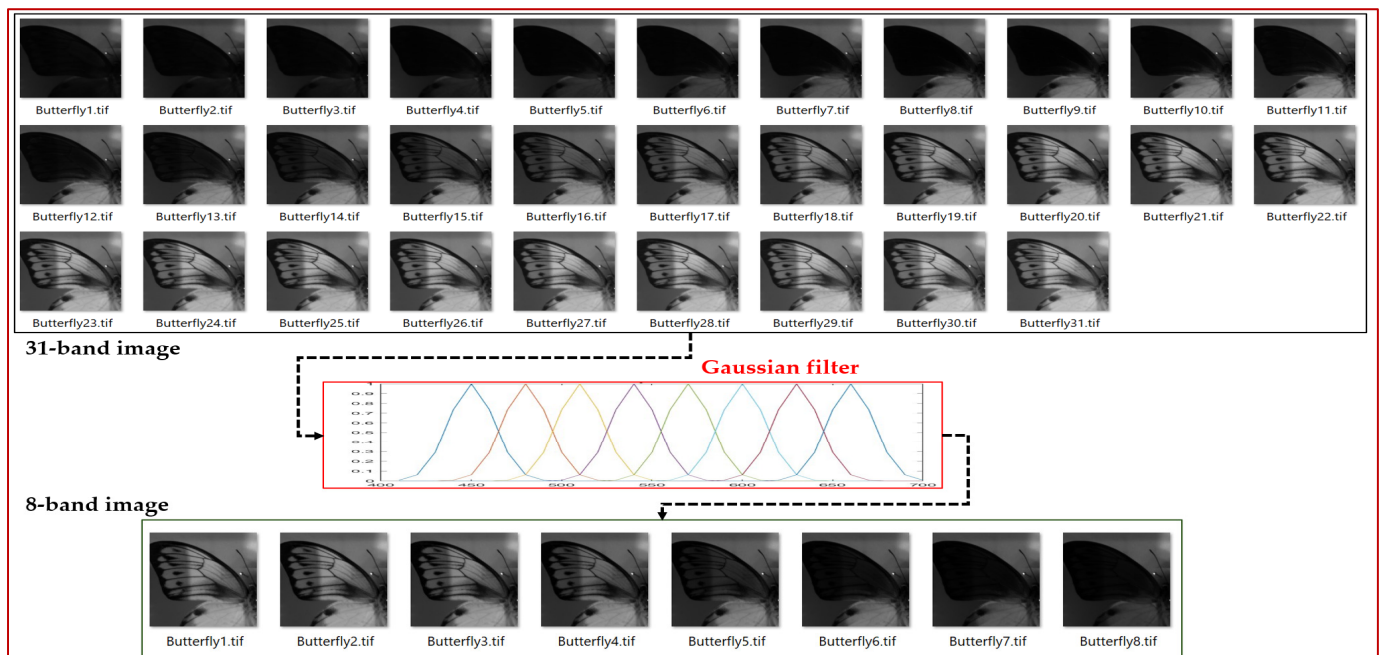


Figure 3. Illustration of the Gaussian projection of images from 31 bands to 8 bands.

Once the Gaussian projection is performed on the 31-band images to reduce them to 8 bands, we move on to the mosaic creation step. This mosaic results from assembling the 8 image strips to form a larger, more informative composite image. This mosaic is then demosaicked using the bilinear method. This bilinear method makes it possible to restore best the images captured by the camera developed as part of the CAVIAR project while preserving spectral details. Using bilinear demosaicking ensures a more faithful representation of the original data, which is essential for subsequent analyses and applications.

Figure 4 shows the images obtained before and after the operation.

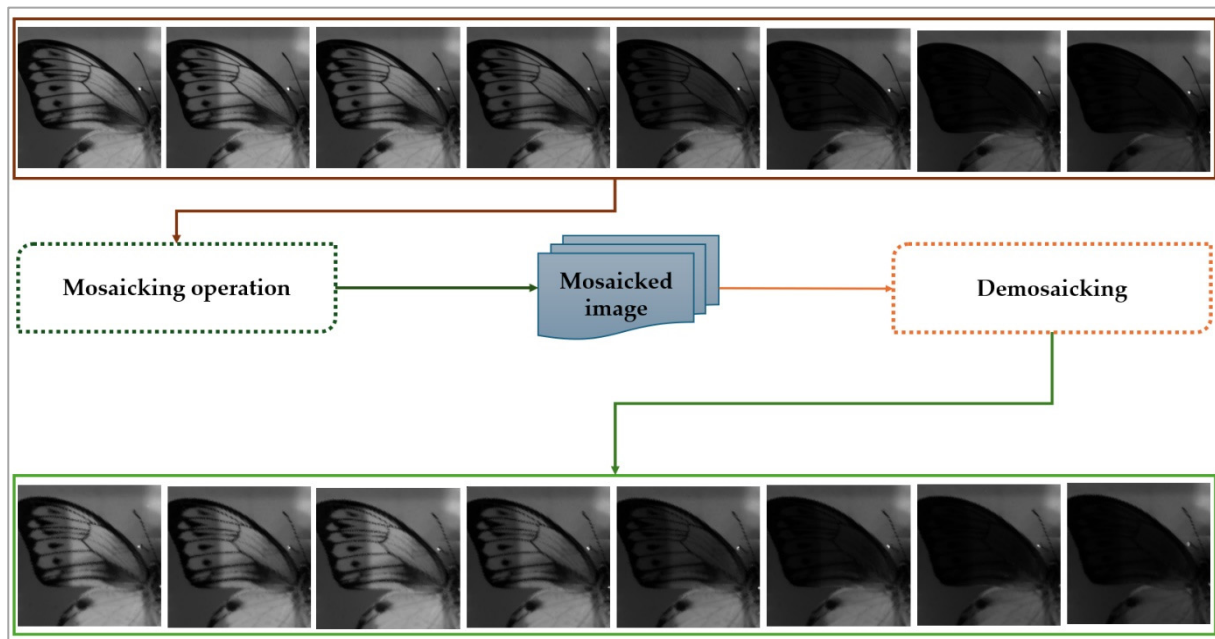


Figure 4. Illustration of the double operation of mosaicking and demosaicking of 8-band images.

Mosaicking is a fundamental process that creates a composite image from multiple-pixel images captured at different wavelengths. This process takes place in several key stages:

1. When capturing an image with our multispectral camera, several individual images are generated. In our case, we obtain eight distinct images, each representing pixels with similar spectral properties. Although they provide detailed information about the observed scene, these images remain incomplete due to their dispersion.
2. The captured pixels are grouped according to their spectral properties. This grouping is essential to reduce data redundancy and improve the clarity of analysis.
3. To reconstruct a fully defined composite image, we rely on our MSFA (multispectral filter array) model [18]. This powerful tool generates mosaics from scattered images while preserving spectral information.

Demosaicking, a crucial part of the process, separates the composite image obtained during mosaicking to find the dispersed images. For each sparse image, demosaicking uses interpolation techniques to estimate the missing pixel values. In our case, we employed bilinear or bicubic interpolation, which evaluates missing values based on neighboring pixels. Once the missing values have been interpolated, the different scattered images are merged to create the demosaicked image.

Figure 6 shows mosaic images from the dataset.

Figure 5 illustrates the mosaicking and demosaicking process.

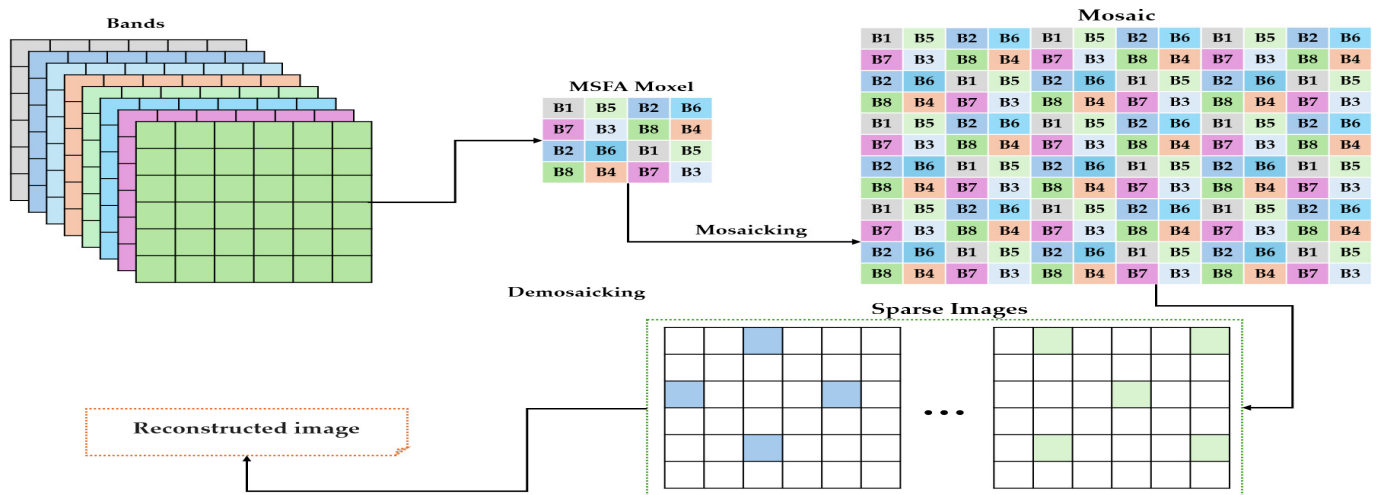


Figure 5. The mosaicking and demosaicking process.

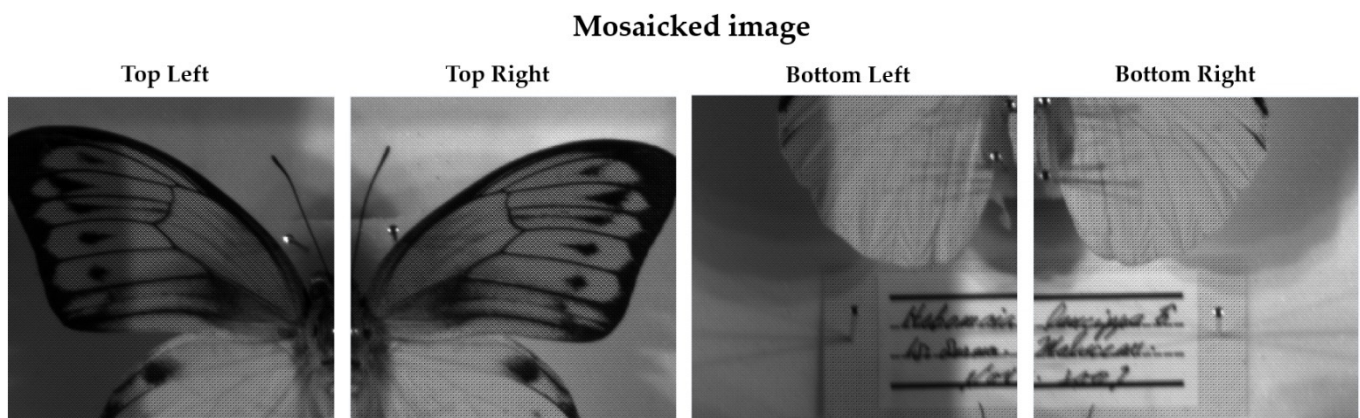


Figure 6. Presentation of mosaicked images.

Figure 7 summarizes the process of preparing our data for the experimentation of our study.

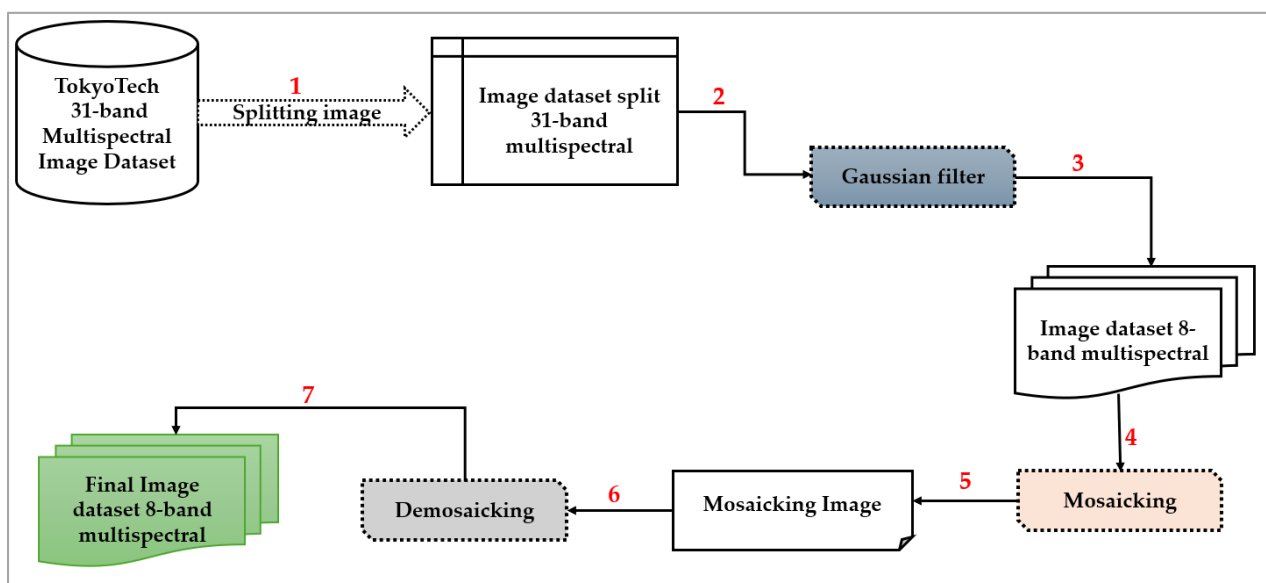


Figure 7. Summary of setting up the dataset.

3.2. Deep-Learning Algorithms

Image reconstruction is a growing field in image processing, aiming to improve image quality with finer and more precise details. Deep-learning algorithms have progressed considerably, enabling more accurate and detailed image reconstructions. Among these algorithms, several innovative approaches stand out for their ability to exploit deep and complex architectures to improve images [19].

In our study, we modify the architecture of three algorithms to meet our specific needs: very deep super-resolution (VDSR), Laplacian pyramid super-resolution network (LapSRN), and deeply recursive convolutional network (DRCN).

Each of these algorithms provides a unique method for tackling the reconstruction problem by exploiting their architectures' specific characteristics to optimize the reconstructed images' quality. When these algorithms process images, they must understand and model how each spectral band contributes to the final image. The spectral response of the cameras plays a role in this process, as it directly influences the quality of the input data for these algorithms. An accurate and well-calibrated spectral response allows algorithms to capture better and restore spectral details, thereby improving the quality and fidelity of reconstructed images.

3.2.1. Very Deep Super-Resolution (VDSR)

VDSR is a sophisticated deep-learning technique designed to improve image resolution. It uses a convolutional neural network (CNN) architecture of 20 layers, significantly more profound than traditional networks. This depth allows VDSR to effectively capture complex features and high-frequency details in images [20].

The architecture includes the following steps:

- **Input processing:** An 8-band multispectral image is fed to the first convolutional layer, which uses 64 filters of size 3×3 and applies a 'same' padding to maintain the dimensions of the input image. A function of ReLU activation is applied to introduce nonlinearity.
- **Hidden layers:** The image output from the first layer will be passed through 19 convolutional layers, followed by a final convolutional layer, which uses several filters equal to the number of output channels and applies 'same' padding. (20th layer). Each convolutional layer uses 64 filters of size 3×3 to extract features from the image.
- **Residual learning:** Another convolution is applied to the input image to bring it into the desired output shape. Residual addition combines the final layer's outputs and the input image's convolution. This will help the model learn the residual details needed to improve the image resolution.

The equation resulting from this model is the following:

- I is the input image.
- K_1, K_2, \dots, K_n is the convolution filters.
- Let R be the result of the last convolution layer.
- Let O be the output image.

$$X_i = \text{ReLU}(K_i * X_{i-1}) \text{ with } X_0 = I \quad (1)$$

$$R = K_{\text{final}} * X_{18} \quad (2)$$

$$F = K_{\text{residual}} * I \quad (3)$$

$$O = R + F \quad (4)$$

Algorithm 1 presents the pseudocode of the VDSR algorithm.

Algorithm 1: VDSRInput: $I, K_1, K_2, \dots, K_n, K_{final}, K_{residual}$ Output: O **Step 1: Initialize the input**

$$X_0 = I$$

Step 2: Apply the sequence of convolutional layersFor $i = 1$ to 18:

$$X_i = \text{ReLU}(K_i * X_{i-1})$$

Step 3: Compute the result of the last convolutional layer

$$R = K_{final} * X_{18}$$

Step 4: Compute the residual connection

$$F = K_{residual} * I$$

Step 5: Calculate the final output image

$$O = R + F$$

Figure 8 shows the VDSR architecture implemented.

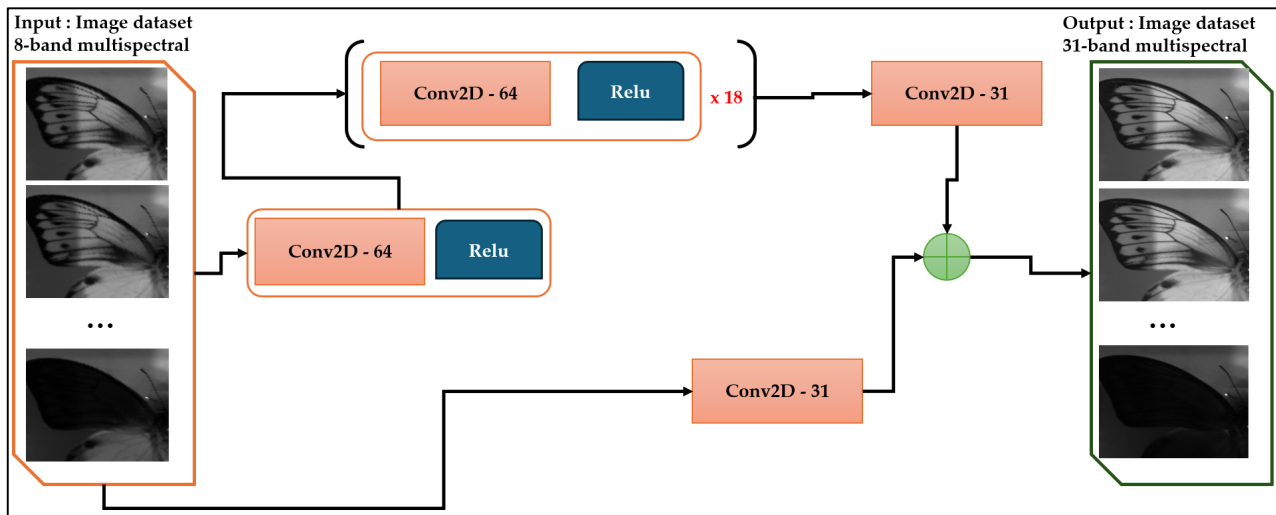


Figure 8. Modified VDSR architecture.

3.2.2. Deeply Recursive Convolutional Network (DRCN)

The deeply recursive convolutional network (DRCN) is an advanced convolutional neural-network architecture for image super-resolution. This model exploits deep recursion to improve the quality of images by increasing their resolution. DRCN is based on applying the same series of convolutional layers recursively, multiple times, to extract and refine image features gradually. This recursion allows the model to be deep and efficient regarding parameters [21]. Its architecture is as follows:

- **Input Layer:** The low-resolution image is first fed into a convolutional input layer, initialing feature extraction.
- **Recursive Convolutional Blocks:** The core of DRCN consists of several convolutional blocks that are applied recursively. Each block includes one or more convolutional layers followed by ReLU activation. The same set of layers is used multiple times to input the image.
- **Recursion:** Instead of directly stacking numerous convolutional layers, DRCN reuses the same layers to obtain an adequate depth without significantly increasing the parameters.
- **Fusion Layer:** Each recursive application's outputs are combined and fused to create a comprehensive, multi-scale image representation. This is a testament to DRCN's design's power to capture the full complexity of the input.
- **Reconstruction Layer:** Finally, a reconstruction layer takes the fused output and generates the final high-resolution image.

The equation resulting from this model is the following:

- **I** is the input image.
- **K** is the convolution filter for feature extraction.
- **K2, K3, ..., Kn** are the recursive layers' convolution filters.
- **K_{final}** is the convolution filter for the reconstruction layer.
- **F** is the extracted feature maps.
- **R_i** is the result of the **i-th** recursive layer.
- **O** is the output image.

$$F = \text{ReLU}(K_1 * I) \quad (5)$$

$$R_i = \text{ReLU}(K_i * R_{i-1}) + F \text{ with } R_0 = F \quad (6)$$

$$O = K_{\text{final}} * R_{16} \quad (7)$$

Algorithm 2 presents the pseudocode of the DRCN algorithm.

Algorithm 2: DRCN

Input: **I**, **K1**, **K2**, **K3**, ..., **Kn**, **K_{final}**

Output: **O**

Step 1: Feature extraction using the first convolutional layer

$$F = \text{ReLU}(K_1 * I)$$

Step 2: Initialize the first recursive layer

$$R_0 = F$$

Step 3: Apply the recursive layers

For **i** = 1 to 16

$$R_i = \text{ReLU}(K_i * R_{i-1}) + F$$

Step 4: Reconstruction layer

$$O = K_{\text{final}} * R_{16}$$

Figure 9 shows the modified DRCN architecture implemented.

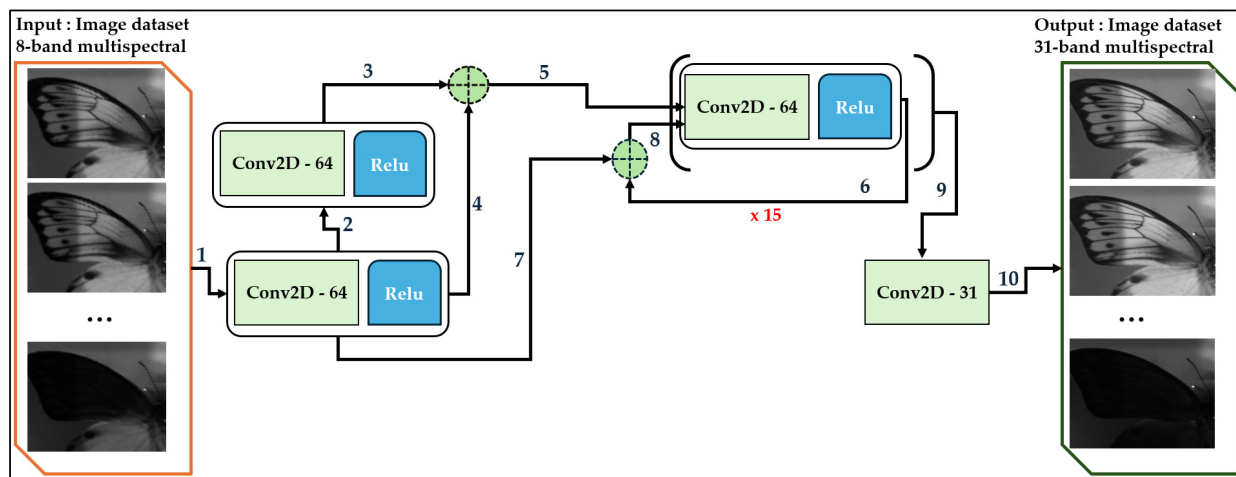


Figure 9. Modified DRCN architecture.

3.2.3. Laplacian Pyramid Super-Resolution Network (LapSRN)

The Laplacian pyramid super-resolution network (LapSRN) is a neural-network architecture designed for image super-resolution. It was introduced to improve image quality and resolution using a hierarchical approach inspired by the Laplacian pyramid.

LapSRN uses a Laplacian pyramid to represent images at different resolution scales. The Laplacian pyramid decomposes an image into a series of frequency subbands, allowing the network to process image details at various resolution levels stepwise [22,23].

- The LapSRN architecture includes several levels, each corresponding to a different resolution scale.
- Each pyramid level contains a convolutional block that extracts image features at that specific scale. This block consists of several convolutional layers followed by ReLU activation.

The equation resulting from this model is the following:

$$F = K_1 * I \quad (8)$$

- F : The initial output of the first transformation applied to the input image.
- K_1 : A filter or convolution kernel applied to the input image.
- I : The input image could be a multispectral or low-resolution image.

This equation represents an initial transformation in which a convolution kernel K_1 is applied to the input image I to obtain the first representation F . This is used to extract low-level features, such as edges or textures.

$$L_j = K_{pyramid} * R_{10} \quad (9)$$

- L_j : An intermediate exit (perhaps at the pyramid level) at step j .
- $K_{pyramid}$: A kernel or filter linked to a pyramid-type structure, often used for multi-scale super-resolutions.
- R_{10} : The final output of the recursive step (probably after 10 iterations).

This equation suggests that the final recursion R_{10} is combined with a specific kernel $K_{pyramid}$, indicating a multi-scale mechanism. This allows information from different levels of detail to be aggregated, thereby enhancing the model's ability to produce images with fine features at different resolutions.

$$O = K_{final} * L_2 \quad (10)$$

- O : The final output of the model.
- K_{final} : A final kernel applied to the previous output for the final transformation.
- L_2 : The output of the second stage in the pyramid or another network module.

This last equation combines the kernel K_{final} with the intermediate output L_2 to produce the result O . This final step can be seen as a refinement, where all the information obtained previously is integrated to generate the final version of the image with improved resolution or other desired properties.

Algorithm 3 presents the pseudocode of the LapSRN algorithm.

Algorithm 3: LapSRN

Input: $I, K_1, K_{r1}, K_{r2}, K_{pyramid}, K_{final}$

Output: O

Step 1: Apply the first convolution to extract features from the input image

$$F = K_1 * I$$

Step 2: Initialize the first residual block

$$R_0 = F$$

Step 3: Apply the residual blocks

For $i = 1$ to 10

$$R_i = K_{r2} * \text{ReLU}(K_{r1} * R_{i-1}) + R_{i-1}$$

Step 4: Create Laplace pyramid levels

For $j = 1$ to 2

$$R_{10} = K_{pyramid} * R_{10}$$

Step 5: Final output

$$O = K_{final} * R_{10}$$

Figure 10 shows the LapSRN architecture implemented.

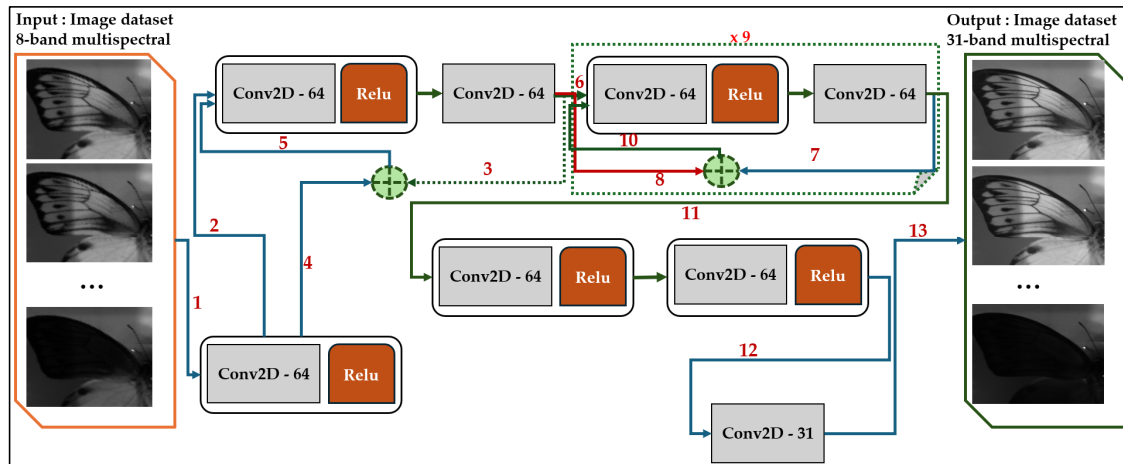


Figure 10. Modified LapSRN architecture.

3.3. The General Process of Our Method

Our general methodology is divided into three main parts:

First, we prepare the data necessary for our study. This step includes collecting, transforming, and normalizing the data to fit our model. We describe how we divide 31-band images into 8-band images using Gaussian projections. These transformed data are then organized into mosaics and demosaicked using a bilinear method, which makes it possible to simulate the images obtained by the camera developed as part of the CAVIAR project.

The second part of our methodology consists of implementing different models based on deep-learning algorithms. We use advanced architectures of algorithms that we have modified to adapt to our study, such as very deep super-resolution (VDSR), Laplacian pyramid super-resolution network (LapSRN), and deeply recursive convolutional network (DRCN). Each model is adjusted to optimize image reconstruction, exploiting the specific characteristics of their respective architectures to improve image resolution and quality. These models are trained with the prepared data, thereby learning to predict the spectral responses of the 31-band images from the 8-band images. Finally, we evaluate our models using test data. This step measures the performance and efficiency of the models. We use various metrics such as loss function, PSNR (peak signal-to-noise ratio), and SSIM (structural similarity index) to evaluate the quality of spectral responses. The evaluation allows us to identify the strengths and limitations of each model and make the necessary adjustments to improve their performance. Figure 11 illustrates the general outline of our methodology.

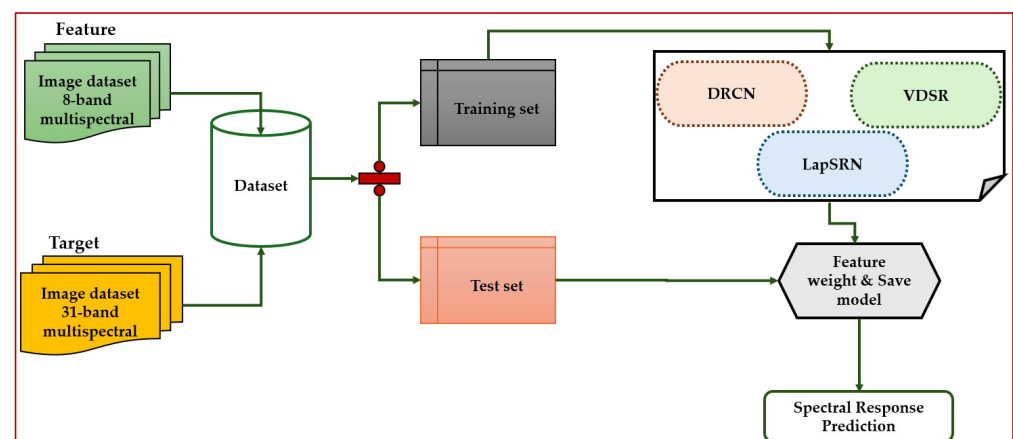


Figure 11. Block diagram of the methodology.

3.4. Metric Performance

To evaluate the performance of the spectral responses provided by our models, we use several metrics [18], including

- PSNR (peak signal-to-noise ratio): Peak signal-to-noise ratio (PSNR) is a commonly used metric to evaluate the quality of a reconstructed or compressed image compared to its original. It quantifies the ratio between the maximum possible power of a signal (the peak signal) and the power of the noise, which alters the fidelity of its representation.

$$PSNR = 10 \cdot \log_{10} \left(\frac{(2^n - 1)^2}{\frac{1}{n} \sum_{i=1}^n (I^i - \hat{I}^i)^2} \right) \quad (11)$$

- The spectral angle mapper (SAM) is a spectral classification method based on calculating the angle between two spectral vectors in a high-dimensional space. Its principle is based on the following elements: Each pixel in a hyperspectral image is represented by a spectral vector, where each dimension corresponds to the reflectance or luminance of the pixel in a specific spectral band. Each class or reference material is also represented by an average spectral vector calculated from selected reference pixels. The SAM calculates the angle between the pixel's spectral vector and each reference vector in n-dimensional space (where n corresponds to the number of spectral bands). The smaller the angle between two vectors, the more similar the spectra are. A maximum angular threshold determines whether a pixel should be assigned to a specific class [24].

$$SAM = \cos^{-1} \left(\frac{\sum_{i=1}^n (I^i \odot \hat{I}^i)}{\sqrt{\sum_{i=1}^n (I^i \odot I^i) \sum_{i=1}^n (\hat{I}^i \odot \hat{I}^i)}} \right) \quad (12)$$

- SSIM (structural similarity index measure): SSIM is designed to assess the similarity between two images by comparing their structural information, luminance, and contrast while considering the characteristics of the human visual system. Unlike more straightforward measures such as mean square error (MSE) or peak signal-to-noise ratio (PSNR), which do not consider aspects of human perception, SSIM offers a more comprehensive assessment of the similarity between the images [25]. The SSIM value is between -1 and 1 , where:

$$SSIM(I, \hat{I}) = \frac{(2\mu_I \mu_{\hat{I}} + C_1)(2\sigma_{I\hat{I}} + C_2)}{(\mu_I^2 + \mu_{\hat{I}}^2 + C_1)(\sigma_I^2 + \sigma_{\hat{I}}^2 + C_2)} \quad (13)$$

- RMSE (root mean square error): RMSE is a metric that evaluates the average error between predicted and actual values in a data set. It is expressed in the same unit as the target value. Here is the RMSE formula:

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (I^i - \hat{I}^i)^2}{n}} \quad (14)$$

- MAE (mean absolute error): MAE is a metric used to measure the accuracy of a model by averaging the absolute errors between the values predicted by the model and the actual values. Unlike RMSE, MAE treats all errors equally without squaring them, making it a more intuitive measure for understanding the average error made by the model. The lower the MAE, the closer the model predictions are to the actual values. The MAE formula is as follows:

$$MAE = \frac{1}{n} \sum_{i=1}^n |I^i - \hat{I}^i| \quad (15)$$

I : Real picture for sample.

\hat{I} : The predicted image for the sample.

4. Experiments and Results

For our experiment, we present the different spectral responses provided by the implemented models obtained from the 8-band MSFA one-shot cameras. We compare our models using the results obtained from our other metrics. The presentation of our results is conducted quantitatively with metrics and qualitatively.

4.1. Experimental Setting

The configuration of the algorithms used for setting up the models is detailed in the Table 1:

Table 1. Summary of algorithm configuration.

Setting	Values
Optimizer	Adam
Learning rate	10^{-3}
Loss function	Mean Squared error
Epochs	100
Batch size	8

4.2. Quantitative Results

Table 2 presents three models suitable for estimating the spectral response of 8-band MSFA one-shot cameras: VDSR, DRCN, and LapSRN, using several performance metrics.

Table 2. Summary of metric values for each model.

Models	Loss	PSNR	SSIM	SAM	RMSE	MAE
VDSR	0.0061	23.6255	0.8135	0.17293	0.06916	0.0525
DRCN	0.0047	25.5059	0.8355	0.13215	0.05849	0.0415
LapSRN	0.0057	23.9505	0.8047	0.17764	0.06534	0.04861

The loss function indicates the error between the predicted and actual images. The values obtained are 0.0061, 0.0047, and 0.0057 for the VDSR, DRCN, and LapSRN models. The DRCN model has the lowest loss, with 0.0047.

PSNR measures the quality of the reconstructed image, with higher values indicating better quality. DRCN achieves the highest PSNR at 25.5059, compared to VDSR at 23.6255 and LapSRN at 23.9505. SSIM evaluates the structural similarity between the original and reconstructed images. Values close to 1 indicate better similarity. DRCN also outperforms other models with an SSIM of 0.8355, while LapSRN has the lowest SSIM at 0.8047.

SAM measures the angle between spectral vectors, with lower values indicating better performance. DRCN has the lowest SAM value, 0.13215, indicating better spectral fidelity than VDSR, 0.17293, and LapSRN, 0.17764.

The RMSE quantifies the average magnitude of errors between the predicted and actual values. A lower RMSE value indicates better model accuracy. DRCN has the lowest RMSE, at 0.05849, while VDSR has the highest, at 0.06916.

MAE, which measures the average absolute errors in predictions, is a critical factor in evaluating the algorithms. DRCN's best performance, with an MAE of 0.0415, highlights its accuracy, while VDSR's highest MAE, at 0.0525, indicates its lower accuracy.

We can list hypotheses for which DRCN performs better:

DRCN (deeply recursive convolutional network) uses deeply recursive architectures that better capture complex and delicate details in images and could justify why DRCN outperforms other models in terms of PSNR, SSIM, and SAM, where accuracy and fidelity to detail are paramount.

In addition, its recursive mechanism would allow predictions to be refined at each iteration, thus capturing more nuances in image reconstruction. This capability could justify its better balance between noise reduction and texture preservation and its superior performance in terms of loss, RMSE, and MAE compared to VDSR and LapSRN models.

Figures 12–14 present the learning curves for the loss, PSNR, and SSIM metrics:

Figure 12 illustrates the learning and validation curves of the VDSR-based model, focusing on three key metrics: PSNR, SSIM, and the loss function (loss). These curves allow one to follow the evolution of the model's performance at each iteration during training.

Figure 13 illustrates the learning and validation curves of the DRCN-based model, focusing on three key metrics: PSNR, SSIM, and the loss function (loss). These curves allow one to follow the evolution of the model's performance at each iteration during training.

Figure 14 illustrates the learning and validation curves of the LapSRN-based model, focusing on three key metrics: PSNR, SSIM, and the loss function (loss). These curves allow one to follow the evolution of the model's performance at each iteration during training.

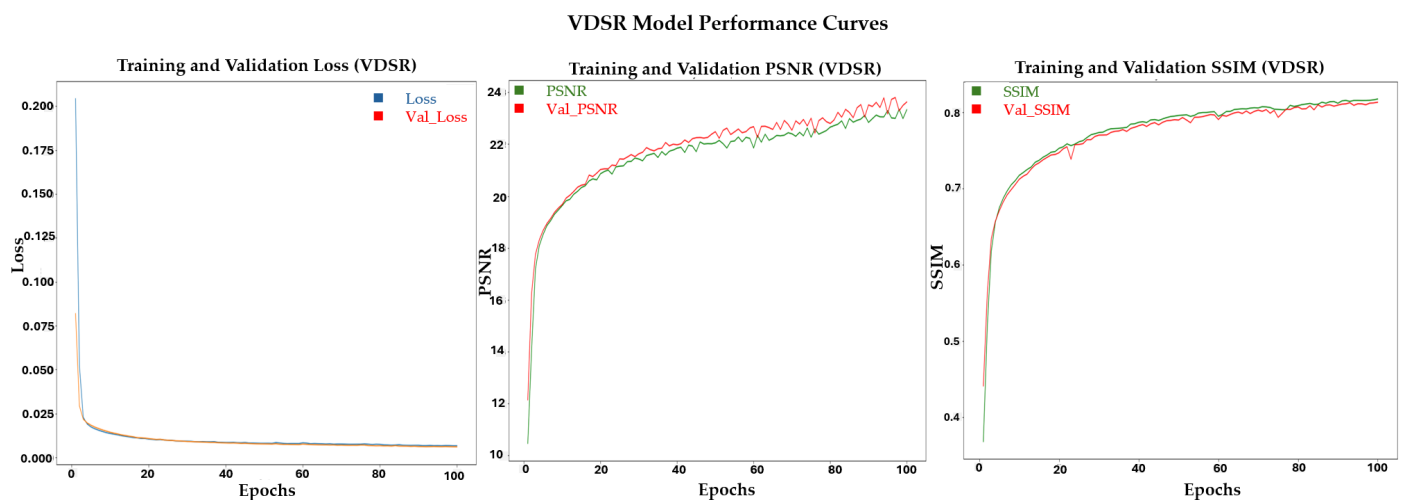


Figure 12. Presentation of the learning and validation curves of the VDSR model for the loss, PSNR, and SSIM metrics.

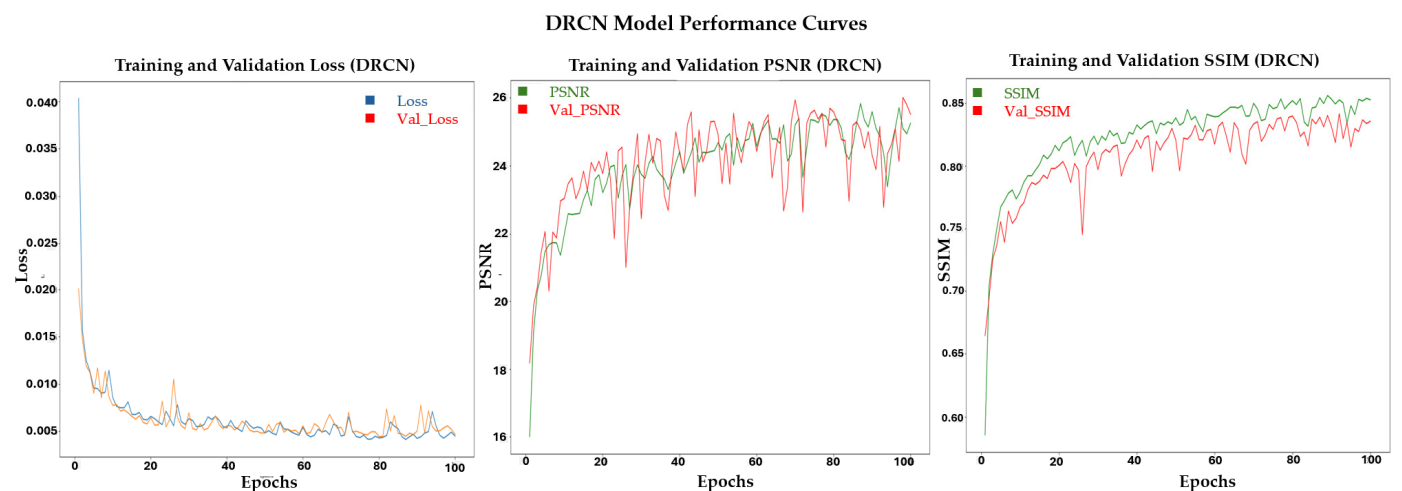


Figure 13. Presentation of the learning and validation curves of the DRCN model for the loss, PSNR, and SSIM metrics.

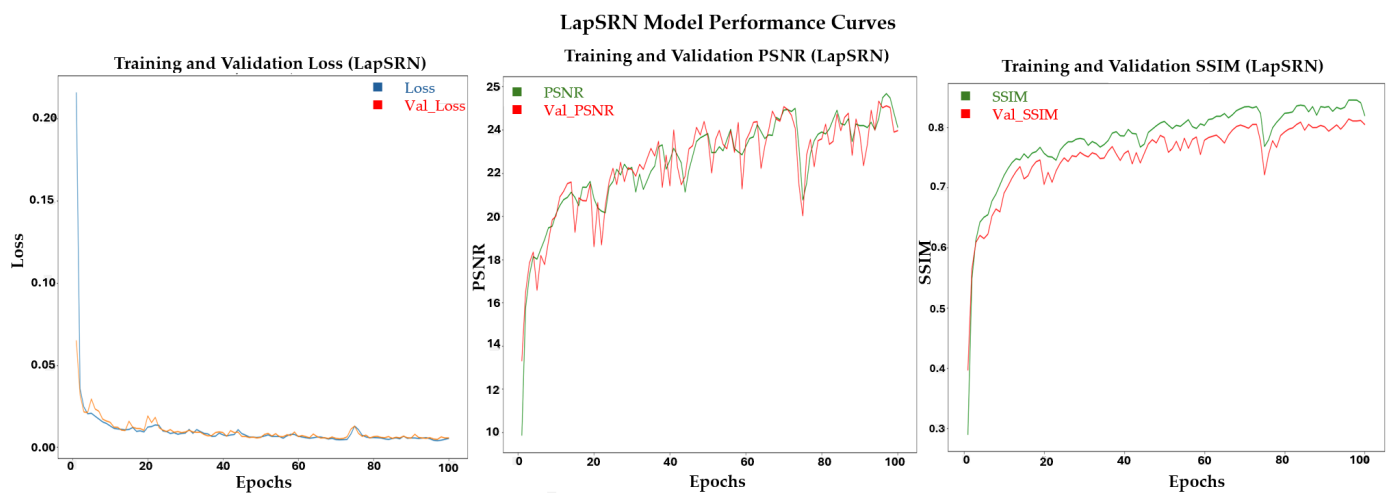


Figure 14. Presentation of the learning and validation curves of the LapSRN model for the loss, PSNR, and SSIM metrics.

4.3. Qualitative Results

This section presents the reference images (ground truth) and the spectral responses from the different models. For clarity, we focus on four specific wavelengths (420, 500, 670, 720) instead of the 31 bands available. This presentation focuses on three categories of images from the TokyoTech database: Chartres, Character, and Tshirts2, selected randomly.

Figures 15–17 show the spectral responses generated by the models and the ground truths.

We observed that the fitted DCRN model provides more accurate spectral responses than the fitted VDSR and LapSRN models, thus confirming previous quantitative results.

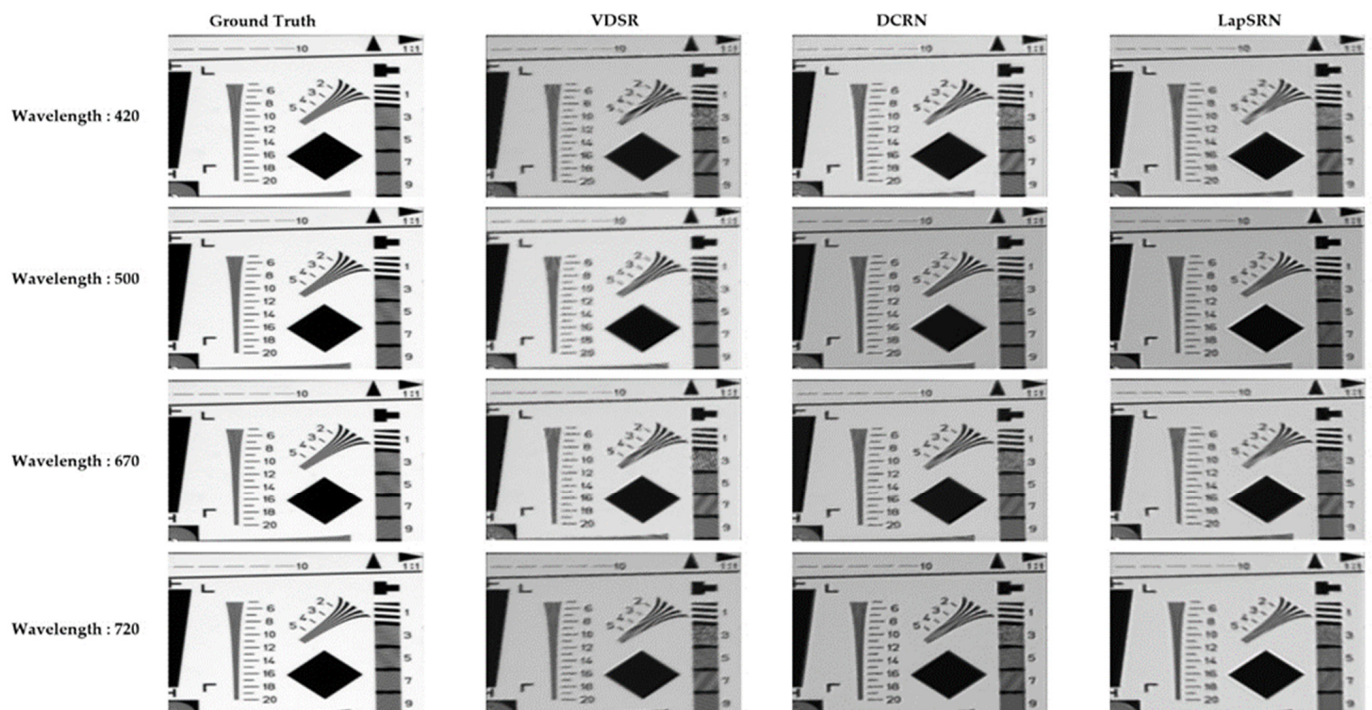


Figure 15. Presentation of the Chartres image, from the ground truth to the spectral response of the adapted VDSR, DCRN, and LapSRN models.

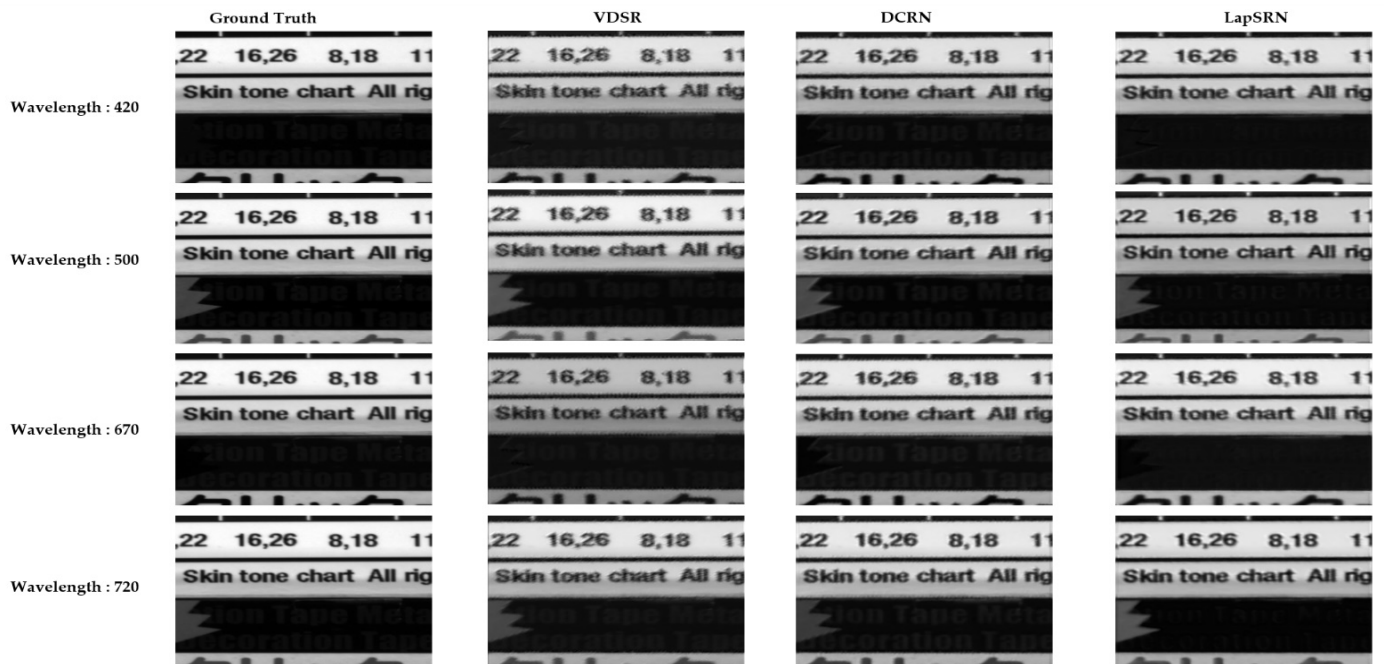


Figure 16. Presentation of the Character image, from the ground truth to the spectral response of the adapted VDSR, DCRN, and LapSRN models.

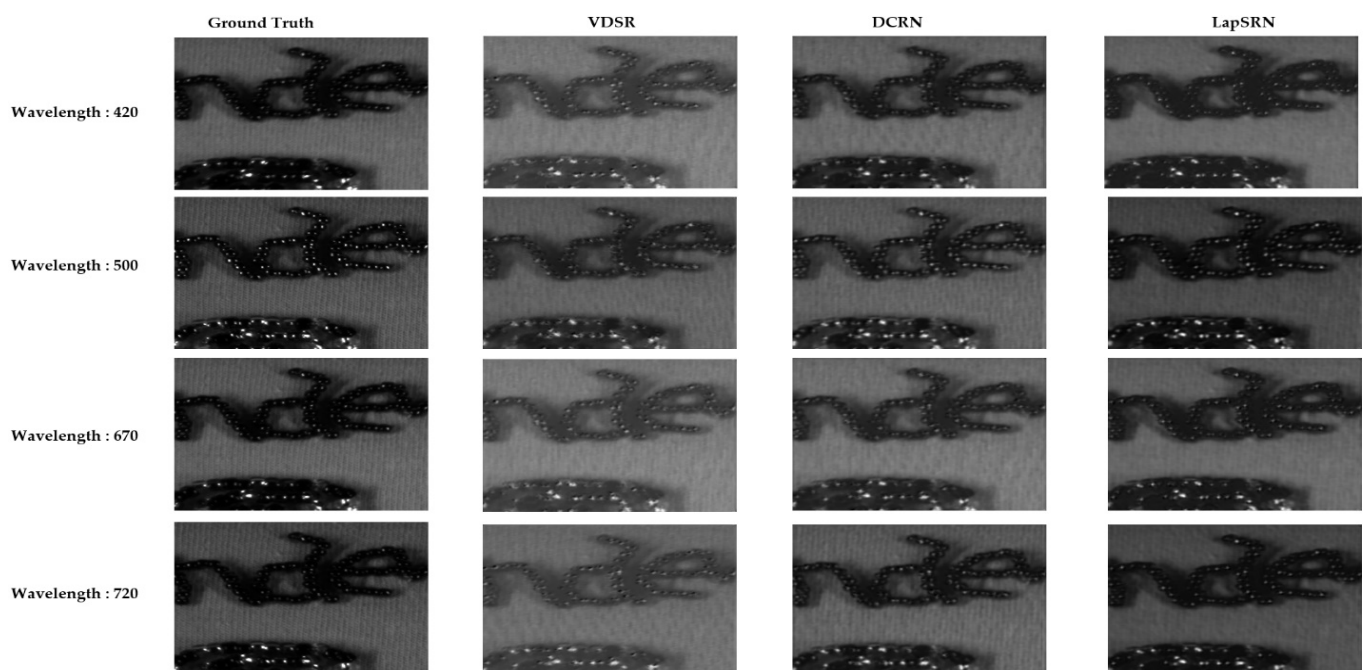


Figure 17. Presentation of the Tshirts2 image, from the ground truth to the spectral response of the adapted VDSR, DCRN, and LapSRN models.

5. Discussion

The performance of the adapted VDSR, DCRN, and LapSRN models reveals significant insights into their capabilities to estimate the spectral responses of 8-band one-shot MSFA cameras. Each model has distinct characteristics that directly influence their effectiveness, and the metric results provide valuable insight into their overall performance.

The DCRN model is the best of the three in quantitative and qualitative assessment. With the lowest values for loss, RMSE (root mean square error), and MAE (mean absolute error), as well as the highest values for PSNR (peak signal-to-noise ratio) and the SSIM

(structural similarity index), DRCN demonstrates an exceptional ability to minimize the error of spectral response estimation. These results show that DRCN does not just reduce the numerical differences between predictions and actual values and generate images perceived to be of better quality than the original images. DRCN's complex architecture enables deep learning of image features, which is the key to superior performance. The depth of the network, combined with advanced optimization techniques, allows DRCN to capture finer details and improve spectral fidelity, which is crucial for applications where color and texture accuracy is paramount.

Although not as efficient as DRCN, the LapSRN model shows interesting results in terms of error minimization (loss, RMSE, MAE) and produces slightly higher-quality images in terms of PSNR compared to VDSR. LapSRN is designed to minimize visual artifacts while maintaining reasonable accuracy, making it suitable for applications where visual quality is more important than structural or spectral fidelity.

VDSR, for its part, stands out for its ability to preserve structures and maintain high spectral precision, as shown by SSIM and SAM (spectral angle mapper) scores. Although its error minimization performance is not as impressive as that of DRCN or LapSRN, VDSR excels in faithfully reproducing details and fine features of images, which is crucial in scenarios where structural fidelity has priority.

Overall, DRCN emerges as the best-performing model for spectral response estimation among those evaluated, providing superior image quality and better spectral fidelity. However, selecting the appropriate model should not be based on overall performance alone; it must also consider the application's specific requirements, including available resources, use conditions, and priorities regarding visual quality, structural fidelity, or spectral precision. DRCN is ideal for applications requiring high accuracy and optimal image quality. At the same time, VDSR and LapSRN could be favored in contexts where other aspects, such as speed or simplicity of the model, are more important.

We use data covering wavelengths from 420 nm to 720 nm with an interval of 10 nm. This setting presents a first limitation: it becomes impossible to reproduce spectral variations lower than this 10 nm interval, which could restrict the precision of spectral reconstructions for applications requiring finer resolution. Furthermore, our results are limited to the visible spectrum since the available data only cover this range. Therefore, extending our findings to other spectral domains, such as infrared, would be infeasible with current data. Future exploration in these areas would require acquiring additional data in the corresponding wavelength ranges. These aspects constitute essential limitations for the interpretation of the results of this study.

Furthermore, the quality of the input data directly influences the results obtained. If the data used are of lower quality, with noise or artifacts, it becomes more difficult for models to accurately capture spectral and structural details. On the other hand, using very high-quality data would significantly improve model performance, leading to more accurate predictions and better spectral fidelity. It is, therefore, crucial to recognize that the size and quality of datasets play a key role in assessing model capabilities.

6. Threats to Validity

It is essential to consider some threats to the validity of this study, which could influence the interpretation of the results.

- The performance of the VDSR, DRCN, and LapSRN models strongly depends on the quality and representativeness of the data used. If the multispectral datasets used for training and validation are not sufficiently diverse or representative of real-world conditions, models could be overfit or underfit, limiting their ability to generalize to other types of cameras or multispectral scenes.
- Model performance also depends on the choice of hyperparameters, such as the networks' learning rates or depths. If these hyperparameters are not rigorously optimized for each model, comparisons may not reflect the maximum capabilities of each architecture.

- The conclusions drawn from this study are based on using a specific multispectral camera (8-band one-shot MSFA). It is important to recognize that these results may not be directly generalizable to other camera configurations, cameras with more or fewer spectral bands, or different application environments. For instance, environments that require more spectral precision or finer spatial resolution may yield different conclusions.

7. Conclusions

Spectral response evaluation of 8-band one-shot MSFA (multispectral filter array) cameras is an innovative research area combining multispectral imaging and advanced deep-learning techniques. In this study, we developed models based on VDSR, DRCN, and LapSRN algorithms, adapted to approximate the spectral response of 8-band one-shot MSFA cameras. We observed significant differences in performance and assessed using key metrics such as loss, PSNR, SSIM, SAM, RMSE, and MAE. The DRCN model performed best, achieving superior results on all metrics considered. Its ability to reduce reconstruction error while preserving the structural details of images makes it a preferred choice for applications requiring high image quality.

On the other hand, the VDSR and LapSRN models, although efficient, show lower performance, which could limit their use in contexts requiring optimal quality. Our study allowed us to develop models to estimate the spectral response of 8-band one-shot MSFA cameras into 31 bands. It is essential to continue this work to improve the quality of the spectral response because the quality of the original images significantly influences the results obtained. The future of image processing lies in developing models that combine performance, efficiency, and adaptability to various situations, ensuring continuous improvement in image quality in different application areas. Future research should focus on improving the input data, optimizing the models used, and exploring other models to find a better balance between performance and efficiency.

For future perspectives, the next step is to validate our approach to the camera developed as part of the CAVIAR project. For this purpose, a database is currently being prepared.

Author Contributions: Conceptualization, K.J.A. and P.G.; methodology, K.J.A. and D.M.; software, K.J.A.; validation, P.G. and D.M.; formal analysis, K.J.A.; investigation, K.J.A. and D.M.; resources, K.J.A. and P.G.; data curation, K.J.A.; writing—original draft preparation, K.J.A. and D.M.; writing—review and editing, P.G.; visualization, D.M.; supervision, P.G. and D.M.; project administration, P.G.; funding acquisition, P.G. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: <http://www.ok.sc.e.titech.ac.jp/res/MSI/MSIdata31.html> (accessed date 14 October 2024).

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Lanaras, C.; Baltsavias, E.; Schindler, K. Hyperspectral Super-Resolution with Spectral Unmixing Constraints. *Remote Sens.* **2017**, *9*, 1196. [CrossRef]
2. Stuart, M.B.; Davies, M.; Hobbs, M.J.; Pering, T.D.; McGonigle, A.J.S.; Willmott, J.R. High-Resolution Hyperspectral Imaging Using Low-Cost Components: Application within Environmental Monitoring Scenarios. *Sensors* **2022**, *22*, 4652. [CrossRef] [PubMed]
3. Cao, S.-Y.; Shen, H.-L.; Chen, S.-J.; Li, C. Boosting Structure Consistency for Multispectral and Multimodal Image Registration. *IEEE Trans. Image Process.* **2020**, *29*, 5147–5162. [CrossRef]
4. Mohammadi, V.; Gouton, P.; Rossé, M.; Katakpe, K.K. Design and Development of Large-Band Dual-MSFA Sensor Camera for Precision Agriculture. *Sensors* **2024**, *24*, 64. [CrossRef] [PubMed]
5. Diaz, N.; Alvarado, A.; Meza, P.; Guzmán, F.; Vera, E. Multispectral Filter Array Design by Optimal Sphere Packing. *IEEE Trans. Image Process.* **2023**, *32*, 3634–3649. [CrossRef] [PubMed]

6. Pu, R. Mapping Tree Species Using Advanced Remote Sensing Technologies: A State-of-the-Art Review and Perspective. *J. Remote Sens.* **2021**, *2021*, 2021–9812624. [CrossRef]
7. Monno, Y.; Kikuchi, S.; Tanaka, M.; Okutomi, M. A Practical One-Shot Multispectral Imaging System Using a Single Image Sensor. *Trans. Image Proc.* **2015**, *24*, 3048–3059. [CrossRef] [PubMed]
8. Toivonen, M.E.; Klami, A. Practical Camera Sensor Spectral Response and Uncertainty Estimation. *J. Imaging* **2020**, *6*, 79. [CrossRef] [PubMed]
9. Darrodi, M.M.; Finlayson, G.; Goodman, T.; Mackiewicz, M. Reference Data Set for Camera Spectral Sensitivity Estimation. *J. Opt. Soc. Am. A* **2015**, *32*, 381. [CrossRef] [PubMed]
10. Han, S.; Matsushita, Y.; Sato, I.; Okabe, T.; Sato, Y. Camera Spectral Sensitivity Estimation from a Single Image under Unknown Illumination by Using Fluorescence. In Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition, Providence, RI, USA, 16–21 June 2012; pp. 805–812.
11. Zhao, H.; Rei, K.; Tan, R.; Ikeuchi, K. Estimating Basis Functions for Spectral Sensitivity of Digital Cameras. In Proceedings of the Image Recognition and Understanding, Kyoto, Japan, 20–22 July 2009.
12. Prasad, D.; Nguyen, R.; Brown, M. Quick Approximation of Camera’s Spectral Response from Casual Lighting. In Proceedings of the IEEE International Conference on Computer Vision Workshops, Sydney, NSW, Australia, 2–8 December 2013; pp. 844–851.
13. Matanga, J.; Gouton, P.; Bourillot, E.; Journaux, S.; Ele, P. Method for the Optimal Approximation of the Spectral Response of Multicomponent Image. *Electron. Imaging* **2019**, *31*, 1–9. [CrossRef]
14. Tian, C.; Zhang, X.; Lin, J.C.-W.; Zuo, W.; Zhang, Y.; Lin, C.-W. Generative Adversarial Networks for Image Super-Resolution: A Survey. *arXiv* **2022**, arXiv:2204.13620.
15. Vizilter, Y.V.; Vygodov, O.V.; Zheltov, S.Y.; Rubis, A.Y. A Way to Interconnect the Multispectral Images for Improved Vision Systems by Using the Methods of Diffusive Morphology. *J. Comput. Syst. Sci. Int.* **2016**, *55*, 598–608. [CrossRef]
16. Andriyanov, N.A.; Vasiliev, K.K.; Dement’ev, V.E. Analysis of the Efficiency of Satellite Image Sequences Filtering. *J. Phys. Conf. Ser.* **2018**, *1096*, 012036. [CrossRef]
17. Andriyanov, N.A.; Vasiliev, K.K.; Dementiev, V.E.; Belyanchikov, A.V. Restoration of Spatially Inhomogeneous Images Based on a Doubly Stochastic Model. *Optoelectron. Instrument. Proc.* **2022**, *58*, 465–471. [CrossRef]
18. Yao, J.Y.A.; Ayikpa, K.J.; Gouton, P.; Kone, T. A Multi-Shot Approach for Spatial Resolution Improvement of Multispectral Images from an MSFA Sensor. *J. Imaging* **2024**, *10*, 140. [CrossRef] [PubMed]
19. Christopher, A.; Hari Kishan, R.; Sudeep, P.V. Image Reconstruction Using Deep Learning. In *Machine Learning Algorithms for Signal and Image Processing*; IEEE: New York, NY, USA, 2023; pp. 65–87. ISBN 978-1-119-86183-6.
20. Kim, J.; Lee, J.K.; Lee, K.M. Accurate Image Super-Resolution Using Very Deep Convolutional Networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 1646–1654.
21. Kim, J.; Lee, J.K.; Lee, K.M. Deeply-Recursive Convolutional Network for Image Super-Resolution. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 1637–1645.
22. Zhang, X.; Song, H.; Zhang, K.; Qiao, J.; Liu, Q. Single Image Super-Resolution with Enhanced Laplacian Pyramid Network via Conditional Generative Adversarial Learning. *Neurocomputing* **2020**, *398*, 531–538. [CrossRef]
23. Lai, W.-S.; Huang, J.-B.; Ahuja, N.; Yang, M.-H. Deep Laplacian Pyramid Networks for Fast and Accurate Super-Resolution. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 5835–5843.
24. Spectral Angle Mapper. Available online: <https://www.nv5geospatialsoftware.com/docs/SpectralAngleMapper.html> (accessed on 13 August 2024).
25. Wang, Z.; Bovik, A.; Sheikh, H. *Structural Similarity Based Image Quality Assessment*. *Digital Video Image Quality and Perceptual Coding*; Series in Signal Processing and Communications; CRC Press: Boca Raton, FL, USA, 2005. [CrossRef]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.