remote sensing

MDPI

Article

Improved Multispectral Target Detection Using Target-Specific Spectral Reconstruction

Nicola Acito [1], Michael Alibani [1] and Marco Diani [2,*]

[1] Department of Information Engineering, University of Pisa, 56122 Pisa, Italy; nicola.acito@unipi.it (N.A.); michael.alibani@unipi.it (M.A.)
[2] Italian Naval Academy, 57127 Leghorn, Italy
* Correspondence: marco.diani@unipi.it

Highlights

What are the main findings?

- Spectral reconstruction can support small-target detection in multispectral imagery when it is adapted to the target of interest.
- Spectral reconstruction models trained without target-aware adaptation may fail to preserve discriminative target features and lead to reduced detection performance.

What are the implications of the main findings?

- The usefulness of spectral reconstruction for downstream applications should be evaluated using task-specific criteria rather than reconstruction accuracy alone.
- Task-aware spectral reconstruction enables more reliable use of multispectral sensors for small-target detection, as shown in UAV-based scenarios.

Abstract

Hyperspectral sensors provide high spectral resolution, enabling accurate material discrimination and effective target detection. However, their practical use is constrained by limited spatial resolution and high acquisition costs. This paper proposes a novel framework to enhance small-target detection in multispectral imagery by leveraging deep learning-based spectral reconstruction to generate high-resolution hyperspectral representations from multispectral inputs. Two state-of-the-art reconstruction networks, MST++ and MIRNet, are trained using paired multispectral–hyperspectral samples derived from AVIRIS-NG data through proper spectral response functions. To improve discriminative capability for the target of interest, a rapid, target-specific fine-tuning stage is introduced, allowing the models to adapt to spectral signatures that are poorly represented or absent in the original training data. Target detection is performed using a spectral signature-based detector applied to the reconstructed hyperspectral data. The proposed framework is evaluated in a real-world scenario involving known field-deployed targets and hyperspectral imagery acquired from an unmanned aerial vehicle. Experimental results demonstrate that the proposed approach significantly outperforms baseline detection applied directly to multispectral data. These findings underscore the effectiveness of spectral reconstruction for downstream tasks such as target detection, particularly in scenarios where hyperspectral data are expensive or unavailable.

Keywords: target detection; spectral reconstruction; hyperspectral; multispectral

Check for updates

Academic Editors: Yuhan Liu, Zhenming Peng, Fei Teng and Xiaoyang Wang

Received: 19 January 2026
Revised: 19 February 2026
Accepted: 28 February 2026
Published: 3 March 2026

Copyright: © 2026 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.

1. Introduction

Hyperspectral (HS) sensors measure the incoming radiation across hundreds of narrow, contiguous spectral bands, providing highly detailed spectral information for effective material differentiation and target detection [1–3]. However, the high spectral resolution often comes at the cost of reduced spatial resolution and significantly higher sensor complexity and operational costs. In contrast, multispectral (MS) sensors, which acquire data in a limited set of broader spectral bands, yield worse spectral discriminability, but get a better spatial resolution and lower costs. Balancing spatial and spectral resolution alongside cost considerations has driven the development of methodologies capable of synthesizing the fine spectral resolution of HS imagery and the high spatial resolution and cost-efficiency of MS data.

In this framework, a promising approach is the one based on spectral reconstruction (SR) techniques [4]. These techniques aim to generate HS-like data from inputs containing only a restricted number of spectral bands thus eliminating the need for costly HS sensors, which are frequently unavailable in many practical applications. Recent advancements in deep learning have led to the development of several SR algorithms capable of reconstructing detailed hyperspectral data from lower-dimensional sources, such as RGB [5,6] and MS images [7–9].

Although SR is an inherently ill-posed problem, these techniques have demonstrated their effectiveness in various downstream tasks, including image segmentation [10], classification [7], and unmixing [11]. However, their impact on target detection (an essential aspect of both military and civilian applications [12]) has received comparatively less attention.

Specifically, this work focuses on the problem of detecting small objects whose spatial extent is on the order of the sensor's ground sampling distance (GSD) and therefore occupy only a few pixels in the image. In such cases, due to the limited number of pixels representing the target, information related to the shape of the object and its texture cannot be effectively exploited for detection, and only spectral information remains relevant. Assuming that the spectral reflectance of the object of interest is available, the problem at hand can be referred to as spectral signature-based target detection (SSBTD) [13].

Notice that it is reasonable to assume that a given SSBTD algorithm yields better performance when applied to hyperspectral data than to multispectral images, since the subtle spectral variations that distinguish targets are better captured. For this reason, in this work we investigate the effectiveness of SR algorithms in addressing the SSBTD problem when only MS images are available. To the best of our knowledge, this problem has not yet been addressed in the literature. In fact, the only study that considers both SR and SSBTD is [14], where the authors presented a method for reconstructing MS images from RGB data with the goal of enhancing the detection of spatially extended and camouflaged targets. However, their research primarily focused on extended targets, rather than on the more complex challenge of detecting small targets, where the phenomenon of spectral mixing of target material with the background is more significant. Furthermore, the work was concerned with the SR of MS images (25 bands from 660 to 950 nm) derived from RGB data.

In the present study, we focus on the detection of small targets by leveraging SR to obtain high–spectral-resolution HS images from MS data in the visible and near-infrared (VNIR) spectral range. For this purpose, we consider off-the-shelf solutions for the SR and SSBTD tasks. Specifically, for the SR task we consider MST++ [15] and MIRNet [16], while for SSBTD we refer to the class of statistically based algorithms [17] and specifically to the widely used normalized matched filter (NMF) [18,19]. We define a training framework based on simulated data for the SR algorithms. A key element of the framework is a lightweight, target-specific fine-tuning (FT) strategy that adapts pre-trained SR models to the spectral signature of the target of interest, thereby improving the reconstruction of

spectral features often underrepresented in standard training data. To address the lack of paired MS–HS data sets, a simulation-driven training strategy based on AVIRIS-NG imagery and realistic sensor spectral response functions (SRFs) is adopted. The proposed methodology is evaluated on a real data set acquired by an HS sensor mounted on an unmanned aerial vehicle (UAV). This data set includes multiple target classes consisting of few-pixel objects, each with known ground truth data in terms of spectral reflectance and spatial location.

Specifically, starting from the UAV acquired HS data, a MS image is generated by applying the SRF of the nine bands of Sentinel 2 (S2), which are closely matched by those of the MAIA S2 camera [20], a commercial MS sensor designed for use on UAVs that replicates the S2 bands in the VNIR spectral range. The MS image obtained serves as input for testing the proposed procedure.

The results indicate that, although generic SR does not necessarily enhance detection performance, the proposed fine-tuned SR approach effectively improves detection accuracy when applied to MS sensors. It is important to emphasize that the focus of this work is neither to propose a new SR method nor to design a novel detector for small targets in MS images. Instead, using a real-world data set and a classical, statistically based detection algorithm, we investigate whether applying standard SR techniques can improve signature-based detection in MS images.

The paper is organised as follows. Section 2 presents the SR-based detection scheme, detailing the AI-based SR algorithms and the adopted detection method. Section 3 describes the training strategy, including a novel fine-tuning procedure designed to enhance detection of specific targets. The data sets used for both training and experimental evaluation are introduced in Section 4, while Section 5 is dedicated to the analysis and discussion of the experiments performed.

2. Proposed Method

2.1. Overview of the Proposed Approach

The goal of this work is to investigate the possibility of using AI-based SR algorithms to improve the performance of the SSBTD task when only MS images are available. The basic processing chain we refer to is illustrated in Figure 1, branch (a). The MS image X serves as input to the SR module whose scope is to obtain a higher spectral resolution image $\hat{H}$. This image is used to detect a specific target whose signature in the spectral domain of $\hat{H}$ is known. This task is accomplished using a specific SSBTD algorithm, which measures the degree of similarity between each image pixel and the target spectral signature.

This general processing pipeline raises several questions that must be tackled. First, the AI-based SR model must be trained using a vast and diverse data set including MS and HS image pairs. This training set is unlikely to be available considering that the goal of the proposed processing chain is to detect targets starting from the MS image and in absence of the HS image. Secondly, in general, target materials are rare and may not be well represented in the training set. This can lead to reconstruction errors for the target of interest, with a direct impact on the final detection performance. In Section 3 we propose original solutions to both the issues raised. In this section, we provide details about the AI-based SR algorithms and the SSBTD strategy used in our experiments.

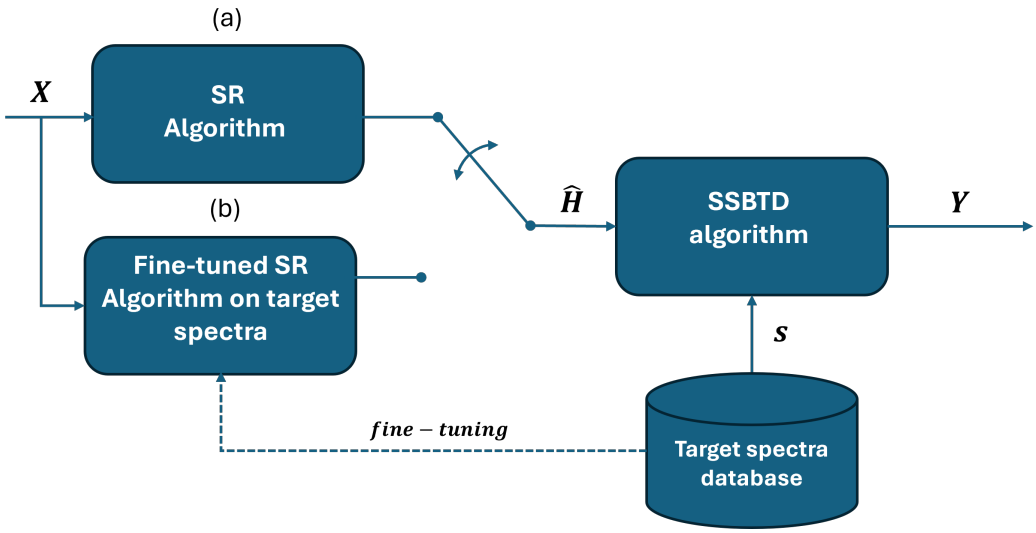

Figure 1. Processing chain of the proposed SR-based detection framework, including the SR block, the SSBTD block, and the target spectral database. (a) Branch using the general SR model without fine-tuning. (b) Branch using the target-specialized SR model with fine-tuning on the target spectrum **s**. The fine-tuning procedure is indicated by the dashed line.

2.2. AI-Based Spectral Reconstruction Algorithms

AI-based SR methods have rapidly gained traction due to their ability to capture intricate spectral patterns directly from data. Several published papers show that SR algorithms based on neural models outperform the prior-based methods such as Wiener estimation [21], pseudo inverse [22,23], and sparse coding [24]. For instance, in [4,25] the authors showed that deep learning methods, specifically those based on the attention mechanism, offer better performance compared to prior-based methods in terms of reconstruction accuracy and adaptability to more practical situations. AI-based SR methods search for a mapping function from the spectral domain of the MS image to that of the HS one:

$$\hat{\mathbf{H}} = f(\mathbf{X}; \mathbf{\Theta}) \tag{1}$$

Equation (1), $\hat{\mathbf{H}} \in \mathbb{R}^{S \times W \times C}$ is the estimated HS image, and $\mathbf{X} \in \mathbb{R}^{S \times W \times c}$ is the MS image. Here, S and W denote the spatial dimension of the images, and C and c denote their spectral dimension (with $c \ll C$). The function $f(\cdot; \mathbf{\Theta})$, depending on the neural model's architecture, is the parametric mapping whose parameters $\mathbf{\Theta}$ are learned during the training process by solving the following optimization problem:

$$\arg\min_{\mathbf{\Theta}} L(f(X; \mathbf{\Theta}), \mathbf{H}) \tag{2}$$

where L indicates the loss function that measures the error between the output of the network and the reference HS image $\mathbf{H} \in \mathbb{R}^{S \times W \times C}$.

For the purposes of this work, we consider two state-of-the-art AI-based SR algorithms, namely MST++ [15] and MIRnet [16]. Originally developed for reconstructing 31-band HS images from three-band RGB inputs, these methods have recently been adapted [9] to obtain high spectral resolution PRISMA HS images starting from Sentinel 2 MS data.

MST++ introduces a transformer framework specifically designed for image reconstruction. Its core innovation is the spectral-wise multi-head self-attention (S-MSA), which forms the basic computational unit called the spectral-wise attention block (SAB). Unlike spatial self-attention methods, S-MSA models correlations across spectral bands in a compact and efficient manner. Several SABs are stacked to form a single-stage spectral-wise transformer (SST), and MST++ cascades multiple SST stages to progressively refine both spatial and spectral representations. The multi-stage design enables MST++ to capture long-

range dependencies while preserving fine spectral details, leading to strong performance and high spectral fidelity.

MIRNet employs a multi-scale residual architecture designed to enhance both local details and global structures in images. Its core innovation lies in the multi-scale residual block (MRB), which combines selective kernel feature fusion (SKFF) with a dual attention unit (DAU). The SKFF module adaptively fuses information across multiple scales, while the DAU refines features via complementary channel and spatial attention. Multiple MRBs are organized into Recursive residual groups (RRGs), forming the primary feature-refinement stages of the network. Through the recursive application of MRBs within each RRG, MIRNet progressively strengthens multi-scale representations, ensuring a balanced integration of fine details and broader contextual cues.

For both MST++ and MIRNet we refer to the PyTorch implementation available at [26]. Of course, MST++ and MIRNet architectures were adapted to the specific case considered in this work in terms of number of bands (c) of the input MS data and number of bands (C) of the HS image.

Specifically, for MST++, the input convolutional layer (embedding block) was modified to accept c-channel MS data and produce C feature maps, while the output convolutional layer (mapping block) was expanded to reconstruct C-band HS images. Accordingly, the internal feature dimensionality was increased to C channels so that all convolutional layers, attention mechanisms, and encoder–decoder components operate in a higher-dimensional spectral space. This enlargement also applies to the global and local residual pathways, whose dimensionality was updated to maintain consistency across all skip connections. In the remaining layers, we retained the configuration of the official implementation ([26]). The number of cascaded SST blocks was set to $N_S = 3$. Within each SST block, the number of SAB modules in the encoder (N_1 and N_2), in the bottleneck (N_3) and in the decoder (N_1 and N_2) was set to 1.

As for MIRNet, the adaptation followed a similar strategy. The input projection layer was redesigned to embed c-band MS data into a C-channel feature space, and the output projection layer was configured to reconstruct C-band HS images. The core components of the network, the MRBs and the RRGs, were configured so that all feature representations operate entirely in C channels, and all residual connections were updated accordingly. For the remaining layers we used the configuration of the original implementation, maintaining the number of RRGs to three ($N_{RRG} = 3$) and using a single MRB (i.e., $N_{MRB} = 1$). These modifications enable MIRNet to accommodate the substantially higher spectral resolution characteristic of HS data.

The loss function L adopted in the training phase of both the networks is the mean relative absolute error (MRAE) [5] as proposed by the original works.

2.3. SSBTD Algorithm

For the target detection task, we adopt the NMF algorithm, which is widely recognized for its effectiveness in detecting both full and sub-pixel targets [18].

The NMF detection score for each pixel is computed as:

$$y_n = \frac{\tilde{\mathbf{s}}^T \tilde{\mathbf{x}}_n}{\| \tilde{\mathbf{s}} \| \| \tilde{\mathbf{x}}_n \|}, \tag{3a}$$

$$\tilde{\mathbf{s}} = \mathbf{C}_b^{-\frac{1}{2}} (\mathbf{s} - \mathbf{m}_b); \quad \tilde{\mathbf{x}}_n = \mathbf{C}_b^{-\frac{1}{2}} (\mathbf{x}_n - \mathbf{m}_b). \tag{3b}$$

In Equation (3a), $\| \cdot \|$ denotes the vector norm, $\mathbf{s} \in \mathbb{R}^C$ is the target spectral signature, and $\mathbf{x}_n \in \mathbb{R}^C$ is the nth image pixel. In Equation (3b), $\mathbf{m}_b \in \mathbb{R}^C$ and $\mathbf{C}_b \in \mathbb{R}^{CxC}$ are the mean vector and the covariance matrix of the background, respectively. According to Equations (3a) and (3b), the detection score y_n for the nth pixel measures the cosine of the

angle between the target spectral signature and the pixel under test in the background whitened space. Based on this geometric interpretation, y_n is in the range $[-1, 1]$ with higher values corresponding to a greater similarity between the target and the pixel under test. Both $\mathbf{C}_b$ and $\mathbf{m}_b$ are estimated from the entire image pixels. The inverse of the covariance matrix is regularized by applying a truncated singular value decomposition inversion ([27]), where eigenvalues smaller than 10^{-9} are set to this threshold.

3. SR Algorithms: Training Strategy

This section describes the solutions proposed to tackle the two issues raised in Section 2.1.

The first issue concerns the fact that training SR models requires a large number of MS–HS image pairs covering various scenarios. The availability of such data set is highly unlikely in practical applications. To deal with this problem, we propose obtaining the necessary data through simulation. The idea is to leverage the availability of a large repository of high-spectral-resolution HS images, along with the nominal SRFs of both the MS and HS sensors, to generate the MS–HS image pairs for training. This is achieved by applying the aforementioned SRFs to each original high-resolution HS image.

From an operational standpoint, this SRF-based mapping can be written in matrix form. Let $\mathbf{H} \in \mathbb{R}^{S \times W \times C}$ denote the HS image cube and $\mathbf{R} \in \mathbb{R}^{C \times c}$ the spectral response matrix of the MS sensor, with columns containing the multispectral SRFs evaluated at the center wavelengths of HS sensor. The simulated MS cube $\mathbf{X} \in \mathbb{R}^{S \times W \times c}$ is obtained by SRF-weighted spectral aggregation applied pixel-wise:

$$\mathbf{x}_{s,w} = \mathbf{h}_{s,w}\mathbf{R} \tag{4}$$

where $\mathbf{x}_{s,w}$ and $\mathbf{h}_{s,w}$ denote the $1 \times c$ and $1 \times C$ vectors associated with the pixel at spatial coordinates (s, w) in the MS and HS images, respectively.

The HS–MS simulation process is sensor-specific and based on the spectral response functions of the considered MS sensor. However, the overall framework is general and directly transferable to other instruments once their SRFs are known.

The freely accessible AVIRIS-NG image repository [28] serves as a highly valuable resource for implementing the proposed strategy. AVIRIS-NG is an airborne HS sensor developed by the Jet Propulsion Laboratory (JPL) that provides high-resolution imagery both spatially and spectrally. AVIRIS-NG covers the spectral range from 380 nm to 2510 nm with a spectral resolution of 5 nm. It captures images with 600 cross-track samples with spatial sampling from 0.3 m to 7 m, and it is ideal for simulating the image pairs required for our work.

Once the MS–HS image pairs combining the AVIRIS-NG data with the available SRFs are obtained, the data set for the SR models' training can be obtained by extracting N image tiles of size $p \times p$ pixels (in our experiments $p = 32$).

The second issue we need to address relates to the potential scarcity of the target spectra in the training set. Targets, especially man-made objects, are unlikely to be present in the AVIRIS-NG data used to simulate the training set. Although the large data set used for general-purpose reconstruction benefits overall model performance, it may not adequately represent rare targets. This could reduce the SR models' ability to effectively reconstruct the target spectra within the proposed processing pipeline, ultimately degrading final detection performance. To deal with this issue, we introduce a fine-tuning procedure aimed at obtaining a specialized SR model for each target of interest. Specifically, starting from the initial SR model trained on the general data set (hereinafter referred to as *original training set*), we propose performing fine-tuning on a smaller data set tailored to the target of interest. This data set is created by randomly selecting $n \ll N$ tile pairs from the *original training set* and injecting the spectral signature of the target. Both the target's position and

size (in pixels) within each MS–HS pair are randomly determined, with the size constrained to cover up to 20% of the tile area. Finally, generated images are processed by a low-pass filter (in the spatial domain) to blend the target pixel with the background.

The block diagram of the proposed procedure including the target specialized (fine-tuned) AI-based SR algorithm is reported in Figure 1, branch (b).

From an operational perspective, the fine-tuning stage assumes that the target of interest is known in advance, which is typical in spectral-signature-based detection scenarios. The fine-tuning process is lightweight compared to full SR model training because it is performed on a reduced set of target-specific samples and can be executed on demand with limited computational cost. Using separate models fine-tuned on demand preserves task-specific spectral features and provides flexibility when the set of relevant targets changes or expands, which is important in real-world applications where new or rare targets may arise. A single model jointly fine-tuned on multiple targets is also feasible; however, this requires the model to represent a broader range of classes, which can dilute discriminative features associated with any particular target and reduce sensitivity to specific spectral signatures. In addition, good performance under one set of targets does not necessarily guarantee similar performance under different or larger sets of classes, and there is no clearly defined number of classes that ensures generalization under a joint training approach. For these reasons, on-demand target-specific fine-tuning remains a practical and adaptable choice.

4. Data Sets: Training Set and Test Set

In this section, we introduce the experimental data set (test set) adopted to assess the performance of the proposed procedure (Section 4.1). Furthermore, we provide details on the training of the SR models and the fine-tuning procedure (Section 4.2).

4.1. Experimental Test Set

For the experimental evaluation of the proposed approach in a real-world scenario, we assembled a test set collected during a dedicated acquisition campaign held in Italy on 25 July 2025. The campaign was carried out using a *Headwall Nano Hyperspec-VNIR* camera mounted on a drone. The camera is a portable instrument that acquires 270 spectral samples within the 400–1000 nm spectral range (VNIR). These spectral samples are generally binned by a factor of 3 to increase the signal-to-noise ratio (SNR) resulting in HS images with $C = 90$ bands, each with full width at half maximum (FWHM) of 6 nm.

During the campaign we executed two flights (with a time interval of nearly half an hour) over the same measurement field in which we deployed different classes of targets across the terrain. The flights were performed at a 27 m altitude, with a GSD of approximately 25 mm. The deployed targets had a size of the order of a few centimeters. The collected HS data are equipped with a ground truth (GT) consisting of the position in the scene and the spectral signature of each target. The latter was obtained by the *ASD FieldSpec 4* instrument that is a handheld spectroradiometer operating in the 350–2500 nm spectral range.

The set of considered targets includes thirteen different object classes, such as plastic toys (of various colors and sizes), a wooden board, plastic plants, and a painted cloth. Each target class is labeled with the letter "T" followed by a number representing its class identifier. For some target classes, multiple objects were placed within the scene. In the remainder of the paper, each individual object within a given class is referred to as "obj" followed by a progressive number (e.g., obj1, obj2, etc.). The different test sets acquired during the two separate flight campaigns are referred to as *Test Set 1* and *Test Set 2* throughout the paper.

Although *Test set 1* and *Test set 2* were acquired using the same sensor and flight configuration, they differ in scene realization. In particular, the positions and orientations of several targets were changed between acquisitions, leading to variations in illumination,

adjacency effects, and background composition. As a result, the two scenes provide distinct detection conditions. Using both data sets increases the number of evaluated target instances and enables a more robust validation on real data.

A complete list of all target classes used in the scene is provided in Appendix A, where, for each class, we report its identification label, a brief description, a photograph taken during the data collection campaign, and the number of objects from that class included in the scene for each flight.

The HS images in *Test set 1* and *Test set 2* have size 640 × 7790 pixels and 640 × 7970 pixels, respectively. Both images contain reflectance data. They were obtained by compensating for atmospheric effects using the empirical line method [29], which exploits the presence of calibration panels (with known spectral reflectance) within the scene. Figure 2 shows the RGB representation of the *Test set 1* HS image where the region of the scene including the deployed targets has been highlighted (red rectangle). Figure 3 shows a zoomed-in view of the region of the scene where the targets are located. In Figure 3 the objects of the same class of targets share the same color. As can be observed, some target classes include more objects. For example, the class named T4 (in black in Figure 3) has nine objects in the scene (see Appendix A for more details).

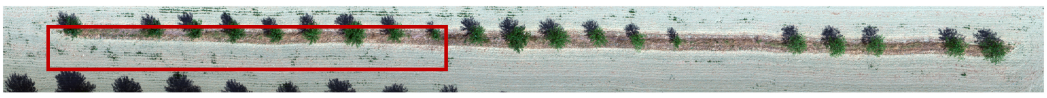

Figure 2. RGB representation of *Test set 1*. The red rectangle indicates the region of the scene including the deployed targets.

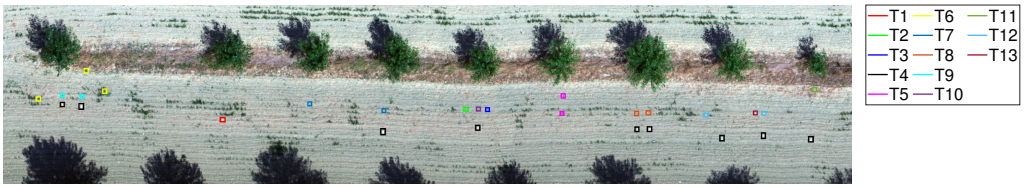

Figure 3. RGB representation of the measurement field of *Test set 1*. In the figure the locations of the targets have been superimposed onto the RGB representation.

The MS images used as input to the proposed procedure are obtained from the 90-band HS images in the two test sets by applying the SRFs of a real MS sensor operating in the VNIR spectral range. We have considered the Sentinel-2 (S2) sensor and, specifically, bands B01–B08 and B08A whose SRFs are made available by the European Space Agency ([30]). Figure 4 shows the SRFs of the nine S2 bands considered. Notably, this configuration closely corresponds to that of the commercial MS camera "MAIA" [20], which is specifically engineered for UAV remote sensing applications.

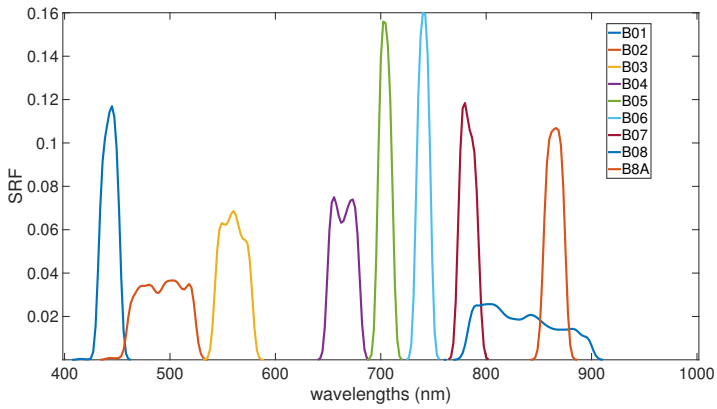

Figure 4. SRFs of the nine S2 channels adopted for simulating the MS images.

4.2. Training Set

Following the training strategy described in Section 3, the *original training set* for the general-purpose SR model was simulated using images acquired by AVIRIS-NG sensors from various regions around the world—including Europe, the USA, and Mexico—spanning the years 2018 to 2022. These data sets encompass a wide range of scenarios, including urban, rural, coastal, and mountainous areas. From the selected AVIRIS-NG data, MS and HS image pairs were generated. The MS images were simulated by applying the SRFs of S2 bands B01–B08 and B08A (Figure 4), while the HS images were obtained using the SRFs of the 90-band HS sensor employed in the measurement campaign. This procedure ensures that the simulated training pairs reproduce the same spectral configurations as those used in the experimental evaluation, namely HS data corresponding to the *Headwall Nano Hyperspec*-VNIR sensor and MS data corresponding to the S2 spectral response functions.

From the resulting data, we extracted $N = 10^4$ tiles of size 32×32 pixels, which constituted the *original training set* for the training of SR models.

According to the procedure described in Section 3, we generated $n = 10^3$ MS–HS tile pairs for each target in our test set to perform the fine-tuning. Specifically, for a given target, its spectral reflectance was injected into n tile pairs randomly selected from the *original training set*. The injections were performed on contiguous pixels, whose position and number are randomly determined.

Training on the *original training set* was performed using an NVIDIA GeForce RTX 4090 GPU (24 GB memory) and the Adam optimizer ($\beta_1 = 0.9$, $\beta_2 = 0.99$) for both MST++ and MIRNet. The initial learning rate was set to $l_r = 10^{-4}$ and was dynamically updated using a cosine annealing schedule. Both networks were trained for 200 epochs using the MRAE loss function, with a total training time of approximately 24 h. For both MST++ and MIRNet, target-specific fine-tuning was performed on the same GPU using the same optimizer and loss function. In the fine-tuning stage, the learning rate was set to $l_r = 10^{-5}$ and kept constant for 10 epochs. All network parameters were updated, with no layer freezing applied. The total fine-tuning time for each network was approximately 10 min, which is significantly shorter than the 24 h required for full training on the *original training set*.

For completeness, the computational complexity of the adopted SR models is summarized in Table 1. Latency was measured as the mean inference time over 50 repeated reconstructions of 128 × 128 images, with batch size equal to one, on the NVIDIA GeForce RTX 4090 GPU used in our experiments.

Table 1. Computational complexity of the adopted SR models (per 128 × 128 input image, batch size = 1). Latency is the mean inference time measured over 50 reconstructions on an NVIDIA GeForce RTX 4090 GPU.

Model	Parameters	FLOPs	Latency
MIRNet	18.98M	182.07G	11.54 ms
MST++	17.78M	125.91G	23.79 ms

5. Experimental Results and Discussion

This section presents the results obtained by applying the proposed procedure to the two test sets described in Section 4.1. Two distinct experiments named *Exp 1* and *Exp 2* are performed. *Exp 1* tests the performance of the procedure depicted in Figure 1 branch (a). Specifically, in *Exp 1* we evaluate, for each target, the detection performance obtained by using the SR models (MST++ and MIRNet) trained on the *original training set* without fine-tuning. *Exp 2* tests the effectiveness of the fine-tuning strategy depicted in Figure 1

branch (b) and described in Section 3. Before presenting and discussing the results of the two experiments in Section 5.1 we introduce the performance metrics adopted.

5.1. Performance Metrics

Performance analysis is carried out using two different types of performance metrics. The first type includes the relative root mean square error (RRMSE) and the spectral angle measure (SAM), which are computed between the HS image obtained by each SR algorithm ($\hat{\mathbf{H}}$) and the reference image ($\mathbf{H}$), provided by the *Headwall Nano Hyperspec-VNIR* camera.

RRMSE and SAM are defined as

$$\text{RRMSE} = \frac{1}{N}\sum_{n}\frac{\parallel \mathbf{H}_n - \hat{\mathbf{H}}_n \parallel}{\parallel \mathbf{H}_n \parallel} \tag{5}$$

$$\text{SAM} = \frac{1}{N}\sum_{n}\arccos\frac{\mathbf{H}_n^T\hat{\mathbf{H}}_n}{\parallel \mathbf{H}_n \parallel \parallel \hat{\mathbf{H}}_n \parallel} \tag{6}$$

In Equations (5) and (6), $\mathbf{H}_n$ and $\hat{\mathbf{H}}_n$ are the nth $C \times 1$ spectral pixels of $\mathbf{H}$ and $\hat{\mathbf{H}}$, respectively. N is the total number of pixels in the images $\mathbf{H}$ and $\hat{\mathbf{H}}$ and ^{T} denotes the vector transpose. Both RRMSE and SAM measure the overall similarity between the original HS image ($\mathbf{H}$) and the reconstructed image $\hat{\mathbf{H}}$.

The second type of performance metric measures detection performance for each object within each target class. In particular, we adopt the false alarm rate (FAR) obtained when detecting 50% of the pixels of the object ($FAR@50\%D$). Denoting as Ω the set of the image pixels not included in the GT (all pixels not labeled as target) and Ω_{obj} the set of image pixels containing the object of interest, FAR for a given threshold value λ is defined as

$$FAR(\lambda) = \frac{\sum_{n\in\Omega} u(y_n - \lambda)}{\text{card}(\Omega)} \tag{7}$$

In Equation (7), according to Equation (3a), y_n denotes the NMF detection score in the nth pixel, $u()$ is the Heaviside step function and $\text{card}(\Omega)$ denotes the cardinality of Ω (i.e., the number of pixels outside the GT). The performance metric $FAR@50\%D$ is obtained as $FAR(\lambda_{50\%})$ with

$$\lambda_{50\%} = \text{median}\{y_n | n \in \Omega_{obj}\} \tag{8}$$

$\lambda_{50\%}$ is the median detection score computed on the pixels of the considered target.

$FAR@50\%D$ is an estimate of the false alarm probability, therefore, lower values indicate better detection performance.

5.2. Results: Exp 1

Here, we refer to the detection scheme of Figure 1 where the SR algorithms trained on the *original training set* are used for all the considered targets. In this case, the overall reconstruction performance achieved by both MST++ and MIRNet on the two test sets is very good (see the first row, labeled as "No FT", of Tables 2 and 3). Specifically, the RRMSE for both SR models is in the order of 0.02, indicating that the error between the $\hat{\mathbf{H}}$ and $\mathbf{H}$ is around 2%. The HS images are also very similar in terms of SAM which attains values lower than 1.8 degrees across both test sets and for both networks. However, the good performance achieved in spectral reconstruction does not correspond to equally good performance in detecting the objects deployed in the scene. This conclusion stems from the results in Tables 4 and 5, which report the $FAR@50\%D$ values obtained for each object within each target class in both *Test set 1* (Table 4) and *Test set 2* (Table 5). The results in the two tables are organized as follows: the first column indicates the target class ID; the second column lists the object ID within each class; and the third, fourth, and fifth columns

show the *FAR*@50%*D* values obtained on MS data, HS data reconstructed by MST++, and HS data produced by MIRNet, respectively. The best results in terms of *FAR*@50%*D* (i.e., the lowest values) are marked in bold. Tables 4 and 5 show that for almost all objects in the two test sets, the detection performance obtained on SR HS images is worse than that obtained on MS data.

The analyzed results confirm the concern raised in Section 2, namely, that without a specific training strategy, the detection task does not benefit from the spectral enrichment provided by AI-based SR algorithms. In fact, it appears that SR even degrades the spectral discrimination capability of the reconstructed data compared to that of the original MS data.

Table 2. *Test set 1*: RRMSE and SAM Values for both MST++ and MIRNet.

Training Strategy	SAM (deg)		RRMSE	
	MST++	MIRNet	MST++	MIRNet
No FT	1.651	1.745	0.023	0.024
T1	1.839	1.734	0.025	0.025
T2	1.870	1.780	0.025	0.025
T3	1.750	1.790	0.025	0.026
T4	1.787	1.809	0.025	0.026
T5	1.802	1.759	0.026	0.025
T6	1.748	1.763	0.025	0.025
T7	1.823	1.712	0.025	0.024
T8	1.786	1.808	0.026	0.026
T9	1.696	1.723	0.024	0.025
T10	1.792	1.784	0.025	0.026
T11	1.830	1.758	0.025	0.025
T12	1.789	1.737	0.025	0.025
T13	1.859	1.724	0.025	0.024

Table 3. *Test set 2*: RRMSE and SAM Values for both MST++ and MIRNet.

Training Strategy	SAM (deg)		RRMSE	
	MST++	MIRNet	MST++	MIRNet
No FT	1.406	1.485	0.023	0.024
T1	1.563	1.518	0.025	0.025
T2	1.539	1.548	0.025	0.025
T3	1.528	1.577	0.025	0.026
T4	1.551	1.579	0.025	0.026
T5	1.558	1.536	0.026	0.025
T6	1.508	1.535	0.025	0.025
T7	1.524	1.466	0.025	0.024
T8	1.564	1.563	0.026	0.026
T9	1.519	1.512	0.024	0.025
T10	1.540	1.561	0.025	0.026
T11	1.515	1.524	0.025	0.025
T12	1.536	1.551	0.025	0.025
T13	1.541	1.466	0.025	0.024

Table 4. *Test set 1*: *FAR*@50%*D* for each object and each target class obtained on the MS image and the HS images reconstructed by MST++ (SR by MST++) and MIRNet (SR by MIRNet).

Tgt ID	Obj ID	MS	SR by MST++	SR by MIRNet
T1	obj1	1.44E-01	1.31E-01	**4.96E-02**
T2	obj1	**1.94E-04**	3.93E-03	1.26E-02
T3	obj1	**1.00E-03**	2.27E-01	6.74E-01

Table 4. *Cont.*

Tgt ID	Obj ID	MS	SR by MST++	SR by MIRNet
T4	obj1	**4.99E-02**	5.06E-01	9.00E-01
	obj2	**7.73E-03**	6.40E-01	8.43E-01
	obj3	**5.57E-02**	8.11E-01	6.98E-01
	obj4	**2.15E-02**	8.65E-01	9.35E-01
	obj5	**9.65E-02**	8.20E-01	4.97E-01
	obj6	**5.93E-02**	9.28E-01	7.45E-01
	obj7	**1.04E-01**	5.68E-01	5.59E-01
	obj8	**6.54E-02**	7.13E-01	7.23E-01
	obj9	**2.69E-02**	5.87E-01	9.41E-01
T5	obj1	**1.16E-02**	4.87E-01	2.36E-02
	obj2	**5.46E-03**	4.01E-01	1.29E-02
T6	obj1	**3.82E-03**	1.20E-01	1.18E-01
	obj2	**2.79E-02**	7.30E-01	3.21E-01
	obj3	**3.33E-04**	2.70E-02	1.54E-01
T7	obj1	**1.06E-04**	7.23E-03	2.49E-01
	obj2	**8.09E-04**	2.55E-02	2.11E-01
T8	obj1	5.84E-04	1.04E-01	**3.57E-04**
	obj2	2.36E-04	6.81E-02	**6.56E-05**
T9	obj1	**3.27E-02**	3.14E-01	9.11E-01
	obj2	**5.01E-02**	4.34E-01	3.15E-01
T10	obj1	**8.13E-04**	3.81E-02	3.12E-01
T11	obj1	**4.69E-05**	1.91E-01	1.07E-01
T12	obj1	4.30E-02	2.08E-01	**4.19E-02**
	obj2	**1.91E-02**	4.48E-01	1.16E-01
T13	obj1	**1.61E-02**	4.94E-01	2.28E-01

Table 5. *Test set 2*: *FAR*@50%*D* for each object and each target class obtained on the MS image and the HS images reconstructed by MST++ (SR by MST++) and MIRNet (SR by MIRNet).

Tgt ID	Obj ID	MS	SR by MST++	SR by MIRNet
T1	obj1	**8.981E-03**	2.120E-02	2.073E-01
T2	obj1	**2.864E-05**	8.287E-03	1.094E-03
T3	obj1	**6.168E-04**	4.243E-01	5.353E-01
T4	obj1	**3.088E-02**	6.342E-01	3.542E-01
	obj2	**1.552E-03**	6.994E-01	5.979E-01
	obj3	**1.350E-02**	9.216E-01	6.160E-01
	obj4	**9.804E-03**	7.474E-01	7.652E-01
	obj5	**1.348E-01**	8.890E-01	4.480E-01
	obj6	**2.510E-02**	9.596E-01	3.621E-01
	obj7	**4.835E-02**	7.883E-01	6.936E-01
	obj8	**1.141E-02**	9.383E-01	5.801E-01
	obj9	**2.120E-02**	8.363E-01	9.549E-01
T5	obj1	**1.906E-03**	3.862E-01	1.874E-02
	obj2	3.057E-02	4.714E-01	**1.098E-02**
T6	obj1	**8.508E-03**	2.828E-01	2.145E-01
	obj2	**1.824E-02**	1.209E-01	2.001E-01
	obj3	**2.525E-02**	5.482E-01	3.656E-01
T7	obj1	**3.040E-05**	3.864E-04	3.341E-03
	obj2	**4.884E-05**	1.544E-03	7.134E-03
T8	obj1	1.830E-04	6.577E-02	**8.904E-05**
	obj2	**1.528E-04**	3.591E-01	4.105E-04
	obj3	**1.700E-05**	2.141E-01	7.812E-04
	obj4	1.204E-03	3.756E-01	**1.032E-03**

Table 5. *Cont.*

Tgt ID	Obj ID	MS	SR by MST++	SR by MIRNet
T9	obj1	**1.697E-02**	3.193E-01	6.739E-01
	obj2	**1.640E-02**	2.326E-01	5.696E-01
T10	obj1	**3.535E-03**	1.355E-01	1.640E-01
T11	obj1	**1.981E-05**	9.892E-02	2.413E-03
T12	obj1	**1.926E-02**	4.054E-02	9.741E-02
	obj2	**7.862E-03**	7.805E-02	1.669E-02
T13	obj1	**5.152E-02**	5.124E-01	6.803E-01

5.3. Results: Exp 2

To show the effectiveness of the proposed procedure—illustrated in the block diagram of Figure 1 branch (b)—this section essentially replicates the performance evaluation of *exp 1* but using for each object in a given target class the fine-tuned SR algorithms specific for that class.

Before discussing target detection results, we would like to point out that fine-tuning does not significantly reduce the overall reconstruction performance of the two considered SR algorithms. This is supported by the results in Tables 2 and 3, which refer to the values of the spectral reconstruction metrics (i.e., RRMSE and SAM) obtained on the two test sets by applying the fine-tuned algorithms. Specifically, with reference to *Test set 1* each row of Table 2 shows the RRMSE and SAM values obtained after applying the versions of MST++ and MIRNet fine-tuned on the target class indicated in the first column. For completeness, the table also includes (in the first row, labelled as "No FT") the values of the performance metrics obtained without fine-tuning. Note that, for all the targets and for both SR algorithms, the RRMSE values remain around 2% while SAM values are below 1.9 degrees. Although the best values for both metrics are achieved without fine-tuning, the degradation observed in the fine-tuned networks is not significant. Similar conclusions can be drawn from the results in Table 3, which presents RRMSE and SAM values for *Test set 2* reported in the same format as Table 2.

As to the detection performance, Tables 6 and 7 show the values of $FAR50\%D$ obtained in the two test sets by involving the fine-tuned SR algorithms. The format adopted is the same as Tables 4 and 5 and the best results are marked in bold. Notice that, $FAR50\%D$ values for MS data are reported here for completeness even though they are the same as those reported in Tables 4 and 5.

Table 6. *Test set 1*: $FAR@50\%D$ for each object and each target class obtained on the MS image and the HS images reconstructed by **fine-tuned** MST++ (SR by MST++) and MIRNet (SR by MIRNet).

Tgt ID	Obj ID	MS	SR by MST++	SR by MIRNet
T1	obj1	1.44E-01	**6.98E-05**	6.02E-03
T2	obj1	**1.94E-04**	**1.85E-05**	4.57E-05
T3	obj1	**1.00E-03**	1.68E-03	8.18E-03
T4	obj1	4.99E-02	**5.12E-03**	1.45E-01
	obj2	7.73E-03	**5.21E-06**	4.49E-05
	obj3	5.57E-02	**9.95E-05**	3.81E-03
	obj4	2.15E-02	**7.20E-05**	4.80E-04
	obj5	9.65E-02	**1.82E-04**	2.22E-04
	obj6	5.93E-02	3.31E-04	**3.29E-04**
	obj7	1.04E-01	**1.53E-03**	8.33E-03
	obj8	6.54E-02	**4.27E-05**	2.02E-03
	obj9	2.69E-02	**1.60E-03**	8.54E-03

Table 6. *Cont.*

Tgt ID	Obj ID	MS	SR by MST++	SR by MIRNet
T5	obj1	1.16E-02	**5.84E-05**	7.82E-05
	obj2	5.46E-03	**4.01E-07**	1.40E-06
T6	obj1	3.82E-03	**0.00E+00**	**0.00E+00**
	obj2	2.79E-02	7.02E-06	**2.01E-07**
	obj3	3.33E-04	**0.00E+00**	**0.00E+00**
T7	obj1	1.06E-04	**1.87E-05**	5.80E-04
	obj2	8.09E-04	1.10E-03	**3.68E-04**
T8	obj1	5.84E-04	**4.01E-06**	1.53E-04
	obj2	2.36E-04	**4.01E-07**	5.42E-05
T9	obj1	3.27E-02	**2.25E-03**	3.01E-03
	obj2	5.01E-02	1.22E-02	**8.39E-03**
T10	obj1	8.13E-04	**0.00E+00**	**0.00E+00**
T11	obj1	4.69E-05	2.21E-06	**0.00E+00**
T12	obj1	4.30E-02	1.00E-05	**2.41E-06**
	obj2	1.91E-02	**4.93E-05**	9.39E-05
T13	obj1	1.61E-02	**3.89E-04**	6.09E-04

Table 7. *Test set 2*: *FAR*@50%*D* for each object and each target class obtained on the MS image and the HS images reconstructed by **fine-tuned** MST++ (SR by MST++) and MIRNet (SR by MIRNet).

Tgt ID	Obj ID	MS	SR by MST++	SR by MIRNet
T1	obj1	8.98E-03	3.33E-06	1.77E-06
T2	obj1	2.86E-05	**1.96E-07**	1.37E-06
T3	obj1	6.17E-04	**6.83E-05**	1.14E-04
T4	obj1	3.09E-02	1.96E-02	**8.83E-03**
	obj2	1.55E-03	**0.00E+00**	1.57E-06
	obj3	1.35E-02	**3.92E-07**	3.77E-05
	obj4	9.80E-03	**1.96E-07**	7.24E-05
	obj5	1.35E-01	**6.10E-05**	1.51E-04
	obj6	2.51E-02	9.55E-05	**3.08E-05**
	obj7	4.83E-02	**7.28E-05**	3.05E-03
	obj8	1.14E-02	**1.26E-05**	4.83E-04
	obj9	2.12E-02	**1.16E-05**	2.46E-03
T5	obj1	1.91E-03	**0.00E+00**	**0.00E+00**
	obj2	3.06E-02	2.39E-05	**1.12E-05**
T6	obj1	8.51E-03	**0.00E+00**	**0.00E+00**
	obj2	1.82E-02	**0.00E+00**	**0.00E+00**
	obj3	2.52E-02	**2.16E-06**	6.21E-04
T7	obj1	3.04E-05	4.81E-05	**1.26E-05**
	obj2	4.88E-05	**1.69E-05**	2.67E-05
T8	obj1	1.83E-04	**2.75E-06**	1.13E-04
	obj2	1.53E-04	**4.31E-06**	5.57E-05
	obj3	1.71E-05	**4.12E-06**	5.10E-04
	obj4	1.20E-03	**3.14E-06**	3.51E-04
T9	obj1	1.70E-02	3.60E-02	**2.78E-03**
	obj2	1.64E-02	**7.98E-05**	1.55E-03
T10	obj1	3.53E-03	7.45E-06	**0.00E+00**
T11	obj1	1.98E-05	**0.00E+00**	**0.00E+00**
T12	obj1	1.93E-02	**0.00E+00**	3.92E-07
	obj2	7.86E-03	**9.81E-07**	5.49E-06
T13	obj1	5.15E-03	**0.00E+00**	1.06E-03

Analyzing the results in Tables 6 and 7, it can be concluded that, for almost all objects, the target detection task benefits from the fine-tuned SR algorithms. In fact, with reference

to *Test set 1* (Table 6), the detection scheme applied to MS data outperforms the others only for obj1 in target class T3, where it attains the lowest $FAR@50\%D$. For the rest of the objects, incorporating SR in the detection scheme significantly enhances performance. To give an example, for obj2 in T4 $FAR@50\%D$ passes from 7.73×10^{-3} to 5.21×10^{-6} when using SR by MST++ and 4.49×10^{-5} in the case of SR by MIRNet. However, we cannot claim that the detection scheme based on either of the two considered SR algorithms consistently outperforms the other.

To better illustrate this point, we introduce the Improvement Factor (IF) index defined as

$$IF = \log_{10} \frac{FAR@50\%D_{MS}}{FAR@50\%D_{SR}} \tag{9}$$

In Equation (9), the subscript MS denotes the $FAR@50\%D$ value obtained on the MS data, whereas SR refers to the $FAR@50\%D$ value attained by the SR-based detection scheme. Thus, for a given object, IF measures—on a logarithmic scale—the ratio between the detection performance obtained on the MS data and that obtained by the SR-based scheme. Positive values of IF indicate that the SR-based scheme outperforms that operating directly on the MS image.

Figure 5 refers to *Test set 1* and, in the first two rows, shows IF values obtained by the SR-based detection schemes employing MST++ and MIRNet, respectively. Specifically, Figure 5 shows IF values for each target, where the label is formed by combining the target class ID with the specific object ID within that class. It is worth noting that, zero values of $FAR@50\%D$ has been replaced by 2×10^{-7} to avoid ill-conditioning in the evaluation of the IF index. $FAR@50\%D = 2 \times 10^{-7}$ corresponds to the presence of a single false alarm pixel in the image.

Figure 5 (first and second rows) shows that for almost all the objects both the detection scheme based on MST++ and MIRNet have positive IF values. Except for obj1 in T3, where both SR-based detection schemes performance worsen, in the remaining cases where IF has negative values, the two SR-based schemes exhibit opposite behavior. Specifically, for obj1 in T4 and obj1 in T7, the MST++ based scheme has a positive IF value in contrast to the MIRNet based scheme. This behavior is reversed for obj2 in T7.

This complementary behavior suggests combining the two SR-based detection schemes in some way, for example, averaging the detection scores produced by the two schemes on a pixel-by-pixel basis. The results of this simple fusion strategy, in terms of IF, are reported in the third row of Figure 5. They show that, except for obj1 in T3, the combination of SR-based schemes actually improves detection performance. This improvement is more significant for some targets. For example, $FAR@50\%D$ decreases by more than three orders of magnitude for class T6.

Results obtained on the *Test set 2* confirm the previous conclusions. Table 7 shows the values of $FAR@50\%D$ obtained on *Test set 2*. It is evident that, for all the targets the best detection performance is achieved by one of the two SR-based schemes. Here-again, for some targets, the two schemes exhibit an opposite behavior. This becomes clear when examining the values of IF presented in Figure 6 for both the MST++ (first row) and MIRNet (second row) based schemes. For obj1 in T7 and obj1 in T9, the scheme based on MIRNet shows positive IF values in contrast to that based on MST++. This behavior is reversed for obj3 in T8. Figure 6 also confirms that combining the two schemes is beneficial in terms of detection performance. Combining the detection scores provided by the two SR-based algorithms actually results in positive IF values for all the targets considered.

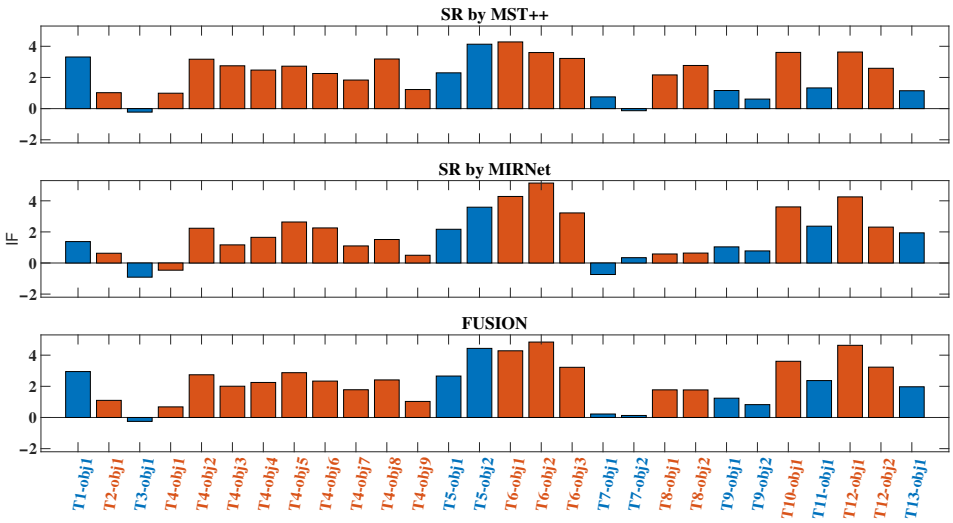

Figure 5. *Test set 1*: values of IF obtained using the SR-based detection framework employing the **NMF** algorithm for the detection task. SR is performed using MST++ (first row, labelled as "SR by MST++") and MIRNet (second row, labelled as "SR by MIRNet"). The third row show results obtained by combining the detection scores provided by the two SR-based detection schemes ("FUSION").

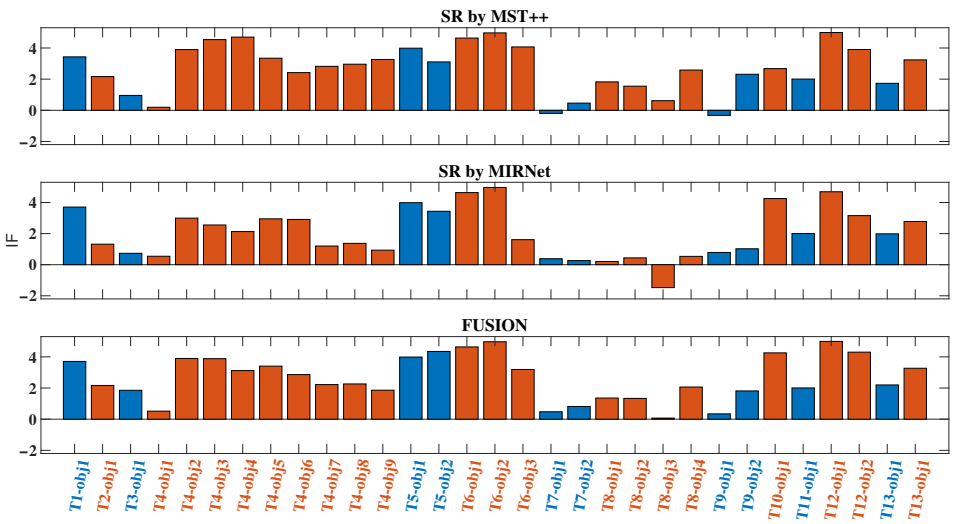

Figure 6. *Test set 2*: values of IF obtained using the SR-based detection framework employing the **NMF** algorithm for the detection task. SR is performed using MST++ (first row, labelled as "SR by MST++") and MIRNet (second row, labelled as "SR by MIRNet"). The third row show results obtained by combining the detection scores provided by the two SR-based detection schemes ("FUSION").

To conclude this section, we present a graphical example of the detection scores obtained on the MS image and by using the SR-based detection schemes. For brevity, we show the results of the fusion strategy, which consists of averaging the detection scores provided by the two SR-based detection schemes (hereinafter referred to as FUSION). The example refers to the *Test set 2* and the target class T6, which contains three objects. This choice is motivated by the fact that T6 is one of the target classes for which we obtained the highest IF. Particularly, starting from a $FAR@50\%D$ in the order of 10^{-2} in the MS case we have obtained $FAR@50\%D$ less than 2×10^{-5} in the FUSION case.

Figure 7a shows the detection score maps obtained from the MS image (first row) and the fusion strategy (third row). For completeness, Figure 7 also includes (second row) the detection score map obtained by applying the fusion strategy to the outputs of the two SR-based detection schemes trained on the *original training set*, i.e., without fine-tuning. In Figure 7a, the objects positions within the scene are superimposed to the detection score maps using red rectangles.

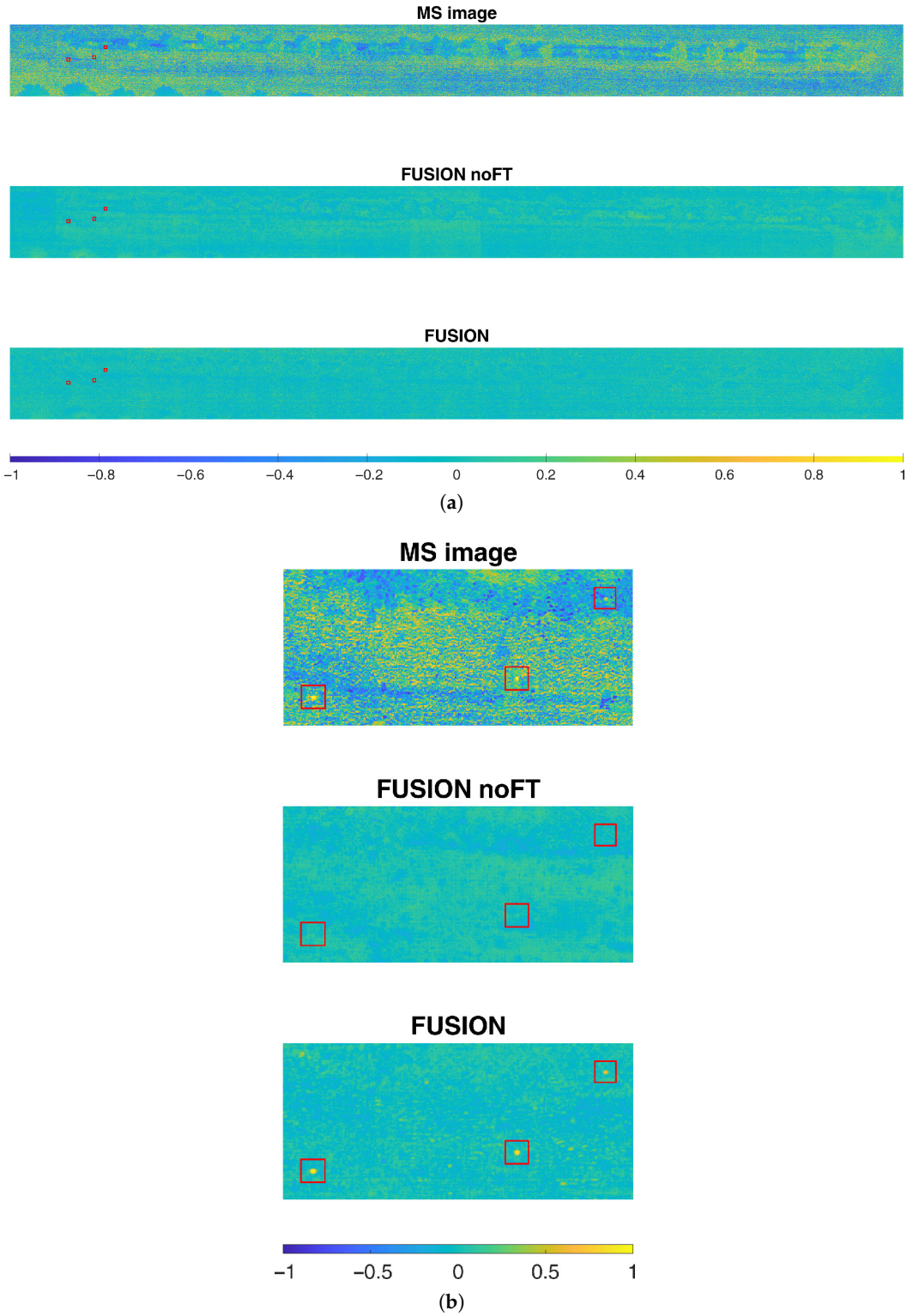

Figure 7. (**a**) *Test set 2*, T6: detection scores obtained on the MS image ("MS image"), averaging the detection scores obtained by both SR-based schemes **without** fine-tuning ("FUSION noFT"), averaging the detection scores obtained by both SR-based schemes **with** fine-tuning. (**b**) zoom-in of the detection scores obtained on the MS image ("MS image"), averaging the detection scores obtained by both SR-based schemes **without** fine-tuning ("FUSION noFT") and averaging the detection scores obtained by both SR-based schemes **with** fine-tuning.

It is evident that the maps obtained by applying the SR-based procedures (both FUSION noFT and FUSION) are smoother than the one derived from the MS image.

Specifically, most pixels in the maps from SR-based procedures have values close to zero in contrast to the MS-based map, which more clearly exhibits texture.

Figure 7b shows zoomed-in versions of the three maps focusing on the area in close proximity to the three objects highlighted by the red rectangles. It is evident that, in the map produced by the fine-tuned SR-based procedures (third row—FUSION), the pixels corresponding to the objects, exhibit high detection scores in contrast to the rest of the scene (background) which shows significantly lower values. Also in the map obtained from the MS data (first row), the pixels containing the objects have high detection scores, although they are not as clearly distinguishable from the background pixels due to the low contrast. The worst case is represented by the map obtained without fine-tuning (second row—FUSION noFT), where the pixels corresponding to the objects exhibit detection scores very similar to those of the background, making them difficult to distinguish.

It is worth highlighting that the FUSION of MST++ and MIRNet outputs is optional and intended to exploit the potential complementarity between different SR architectures. Averaging was chosen as the fusion strategy because it is parameter-free and reproducible, allowing a direct assessment of complementarity without introducing additional learned fusion models. This preliminary, parameter-free evaluation provides an initial practical way to leverage architectural differences and improve performance without assuming or proving a formal theoretical complementarity and can be followed by more advanced fusion strategies in future work.

6. Conclusions

In this paper, we introduce a novel operational workflow to enhance small-target detection in MS data by leveraging spectral reconstruction and target detection techniques. We propose two SR-based detection schemes that share the same detection method (NMF) but use two AI-based SR algorithms: MST++ and MIRNet. For the SR models, we adopt a learning strategy based on simulated data that incorporates the specific SRFs of both MS and HS sensors. The strategy includes two stages: a general-purpose SR model training followed by target-specific fine-tuning. The fine-tuning step is performed on a reduced training set and requires significantly less time than training the general-purpose model.

Experiments on two HS UAV-acquired test sets are presented. The results in terms of $FAR@50\%D$ show that SR-based schemes without fine-tuning are generally not effective, while fine-tuning significantly improves detection performance for most targets.

These results support the potential of the proposed approach. The study is not intended as an exhaustive evaluation of SR effectiveness in target detection, but as a foundation for a detection-oriented framework for MS images. Future investigations with additional data sets, including challenging visibility conditions such as fog, are expected to provide further insights.

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

Funding: Work partially supported by the Italian Ministry of Education and Research (MUR) in the framework of the FoReLab project (Departments of Excellence).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available on request from the corresponding author. The data are not publicly available as they are part of an ongoing research project and can only be shared upon reasonable request.

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

Abbreviations

The following abbreviations are used in this manuscript:

ACE	Adaptive Cosine Estimator
ADAM	Adaptive Moment Estimation
AI	Artificial Intelligence
AMF	Adaptive Matched Filter
AVIRIS	Airborne Visible/Infrared Imaging Spectrometer
AVIRIS-NG	Airborne Visible/Infrared Imaging Spectrometer-Next Generation
CEM	Constrained Energy Minimization
DAU	Dual Attention Unit
ESA	European Space Agency
FAR	False Alarm Rate
FOV	Field of View
FT	Fine-Tuning
FWHM	Full Width at Half Maximum
GPU	Graphics Processing Unit
GSD	Ground Sampling Distance
GT	Ground Truth
HS	Hyperspectral
IF	Improvement Factor
JPL	Jet Propulsion Laboratory
MAIA	MAIA Multispectral UAV Camera
MIRNet	Multi-scale Image Restoration Network
MRAE	Mean Relative Absolute Error
MS	Multispectral
MRB	Multi-scale Residual Block
MST++	Multi-Stage Spectral-wise Transformer++
NMF	Normalized Matched Filter
PRISMA	PRecursore IperSpettrale della Missione Applicativa
RGB	Red, Green, and Blue
RRG	Recursive Residual Group
RRMSE	Relative Root Mean Square Error
S2	Sentinel-2
SAB	Spectral-wise Attention Block
SAM	Spectral Angle Measure
SKFF	Selective Kernel Feature Fusion
S-MSA	Spectral-wise Multi-head Self-Attention
SNR	Signal-to-Noise Ratio
SR	Spectral Reconstruction
SRF	Spectral Response Function
SSBTD	Spectral Signature-Based Target Detection
SST	Single-stage Spectral-wise Transformer
SVD	Singular Value Decomposition
UAV	Unmanned Aerial Vehicle
VNIR	Visible and Near-Infrared

Appendix A

List of the Targets in the Test Sets

This appendix provides a detailed overview of the targets used in *Test set 1* and *Test set 2*. The selection of these targets was intended to cover an assorted range of object in terms of types, materials, colors, and sizes, allowing for a comprehensive evaluation of the models under diverse spectral and spatial conditions.

Table A1 presents all the targets considered in the experiments. For each target, the table includes the target class ID (TgT ID), the number of distinct objects within that target class (N Obj) present in the scenes, a representative picture of the target class, and a brief description. The description specifies the color, the type of material, and the approximate dimensions of the object.

Table A1. Description of the objects deployed in *Test set 1 (T.s. 1)* and *Test set 2 (T.s. 2)*.

Tgt ID	N Obj *(T.s. 1/T.s. 2)*	Picture	Description
T1	1/1		Small airplane model made of yellow plastic, approximately 20 cm wingspan.
T2	1/1		Small tank model made of green plastic, approximately 9 × 4 cm in size.
T3	1/1		Small tank model made of grey plastic, approximately 9 × 4 cm in size.
T4	9/9		Flat wooden boards, approximately 16 × 6 cm in size.
T5	2/2		Small Lego models made of brown plastic, approximately 10 × 13 cm in size.
T6	3/3		Small fake plants made of green plastic, approximately 9 cm in diameter.
T7	2/2		Groups of toy soldiers made of grey plastic, approximately 20 × 15 cm in size.
T8	2/4		Groups of toy soldiers made of red plastic, approximately 20 × 15 cm in size.
T9	2/2		Groups of toy soldiers made of green plastic, approximately 20 × 15 cm in size.
T10	1/1		Small vehicle model made of green plastic, approximately 10 × 3.5 cm in size.
T11	1/1		Synthetic wool towel painted to match terrain color, approximately 16 × 21cm in size.
T12	2/2		Wooden boards painted to match terrain color, approximately 13.5 × 6 cm in size.
T13	1/1		Small plastic animal painted to match terrain color, approximately 12 × 2 cm in size.

All targets were manually annotated across both experiments, ensuring a reliable basis for quantitative comparisons and object-specific performance analysis.

Appendix B

Additional Experiments

It is worth nothing that all the results presented in the previous sections were obtained using the NMF algorithm, which is closely related to the Adaptive Cosine Estimator (ACE), the detection score of the latter being the square of that of the NMF. The choice of NMF is motivated by the fact that it has been shown to be more effective on multi-dimensional data (both MS and HS) than other SSBTD algorithms, despite the discrepancy between the model adopted in their derivation and the physical reality [18].

For completeness, in this section we show some additional results obtained using two different statistical-based SSBTD algorithms, namely, adaptive matched filter (AMF, [31]) and constrained energy minimization (CEM, [32]).

The AMF detection score for each pixel is computed as

$$y_n = \frac{\tilde{\mathbf{s}}^T \tilde{\mathbf{x}}_n}{\| \tilde{\mathbf{s}} \|} \tag{A1}$$

while, the CEM detection score for each pixel is defined as

$$y_n = \frac{\mathbf{s}^T \mathbf{R_b}^{-1} \mathbf{x}_n}{\mathbf{s}^T \mathbf{R_b}^{-1} \mathbf{s}} \tag{A2}$$

In Equation (A1) and Equation (A2), $\tilde{\mathbf{s}}$, $\tilde{\mathbf{x}}_n$, $\mathbf{s}$ and $\mathbf{x}_n$ are the same as those in Section 2.3, while $\mathbf{R_b}$ is the correlation matrix of the background estimated globally by considering all the pixels of the analyzed image.

To limit the length of the paper, we present the results obtained by both the AMF (Figure A1) and CEM (Figure A2) algorithms on *Test set 1* in terms of the IF metric, using the versions of the SR algorithms after FT. It should be noted that these figures do not include the results obtained through the fusion strategy, which are instead presented in Figure A3 (*Test Set 1*) and Figure A4 (*Test Set 2*). In particular, Figures A3 and A4 display, for comparative purposes, the IF values resulting from the fusion of the detection scores provided by the two SR-based detection schemes when the detection task is performed using NMF (first row), AMF (second row), and CEM (third row).

Figure A1 shows that, when using AMF as the detection algorithm, the detection scheme based on MST++ yields positive values of IF for almost all considered objects, except for obj1 in T3 and obj1 in T4. Results in terms of IF are slightly worse when MIRNet is adopted. However, by comparing the results in Figure A1 with those obtained using the NMF detection algorithm (Figure 5), it is evident that the NMF-based scheme outperforms the AMF-based one.

Similarly, when using the CEM detector, the schemes based on MST++ and MIRNet yield positive values of IF (Figure A2) for almost all the considered objects, confirming the effectiveness of SR applied to SSBTD in MS imagery.

Figure A3 shows that, after applying the straightforward fusion strategy, which consists of averaging the detection scores provided by the SR schemes based on MST++ and MIRNet, the best results are obtained using NMF and CEM, with NMF performing slightly better for all target classes except for T4.

It is worth noting that, although not reported here for brevity, the results consistently confirm that the SR-based detection scheme is ineffective without FT, regardless of the detection algorithm employed.

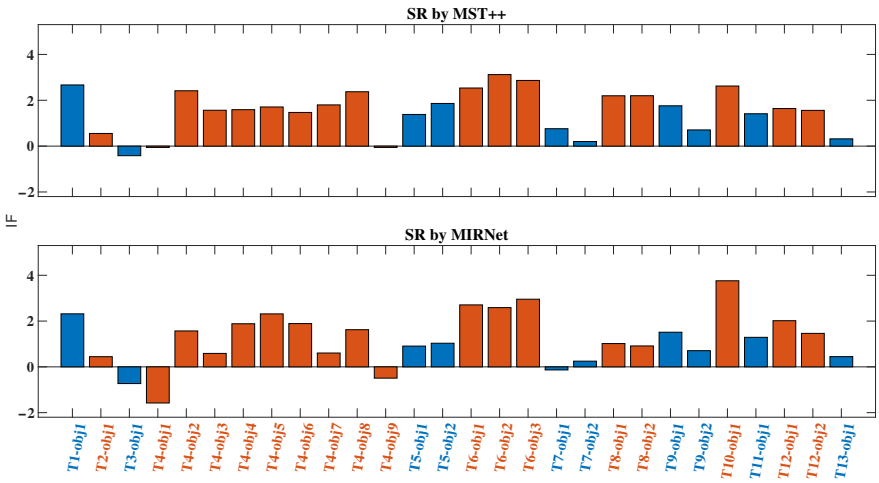

Figure A1. *Test set 1*: values of *IF* obtained using the SR-based detection framework employing the **AMF** algorithm for the detection task. SR is performed using MST++ (first row, labeled as "SR by MST++") and MIRNet (second row, labeled as "SR by MIRNet"). The third row show results obtained by combining the detection scores provided by the two SR-based detection schemes ("FUSION").

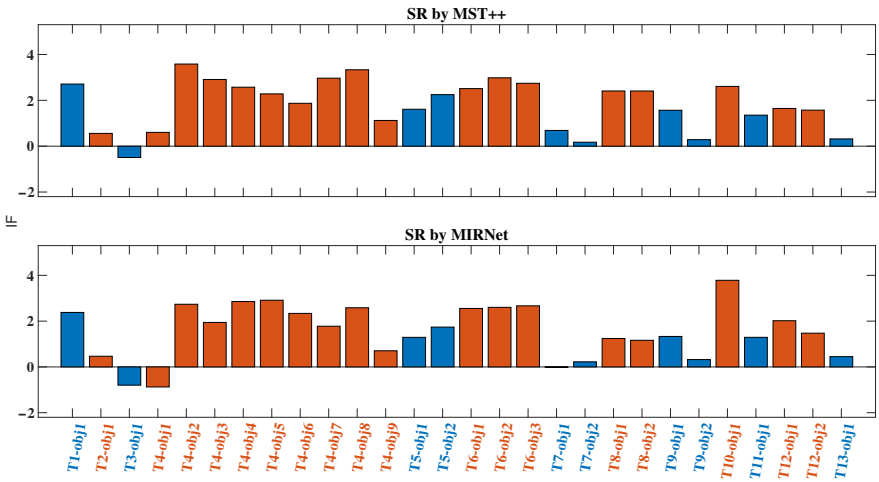

Figure A2. *Test set 1*: values of *IF* obtained using the SR-based detection framework employing the **CEM** algorithm for the detection task. SR is performed using MST++ (first row, labeled as "SR by MST++") and MIRNet (second row, labeled as "SR by MIRNet"). The third row show results obtained by combining the detection scores provided by the two SR-based detection schemes ("FUSION").

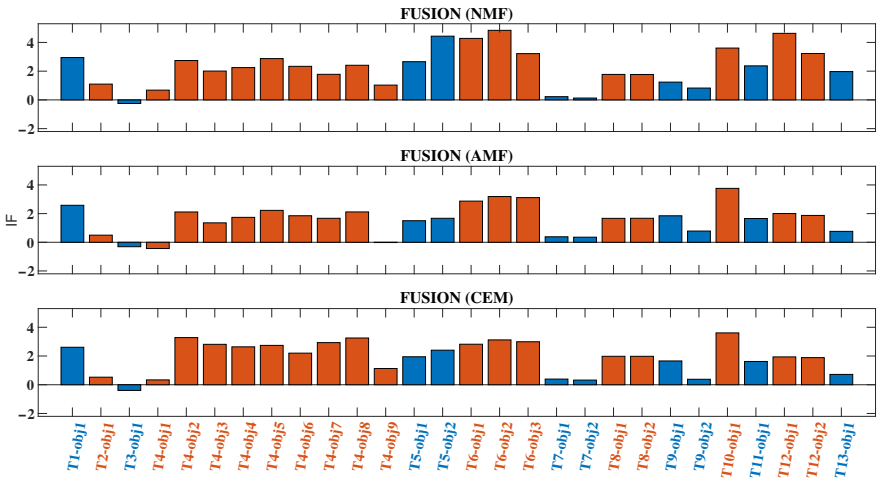

Figure A3. *Test set 1*: values of *IF* obtained by combining the detection scores provided by the two SR-based detection schemes when the detection task is performed by NMF (first row), AMF (second row), CEM (third row).

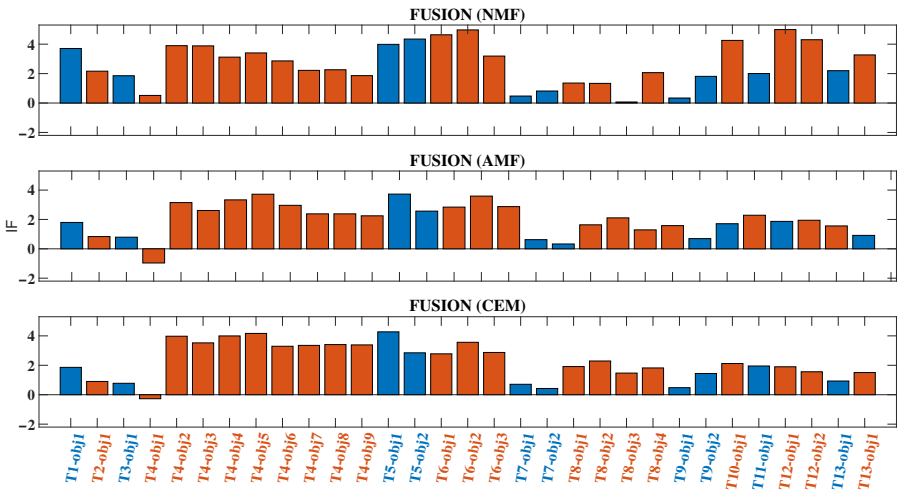

Figure A4. *Test set 2*: values of *IF* obtained by combining the detection scores provided by the two SR-based detection schemes when the detection task is performed by NMF (first row), AMF (second row), CEM (third row).

References

1. Manolakis, D.G.; Lockwood, R.B.; Cooley, T.W. *Hyperspectral Imaging Remote Sensing: Physics, Sensors, and Algorithms*; Cambridge University Press: Cambridge, UK, 2016.
2. Bedini, E. The use of hyperspectral remote sensing for mineral exploration: A review. *J. Hyperspectral Remote Sens.* **2017**, *7*, 189–211. [CrossRef]
3. Acito, N.; Diani, M.; Alibani, M.; Corsini, G. Matched Filter Based on the Radiative Transfer Model for CO2 Estimation From PRISMA Hyperspectral Data. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 1–13. [CrossRef]
4. Zhang, J.; Su, R.; Fu, Q.; Ren, W.; Heide, F.; Nie, Y. A survey on computational spectral reconstruction methods from RGB to hyperspectral imaging. *Sci. Rep.* **2022**, *12*, 11905. [CrossRef] [PubMed]
5. Arad, B.; Timofte, R.; Yahel, R.; Morag, N.; Bernat, A.; Cai, Y.; Lin, J.; Lin, Z.; Wang, H.; Zhang, Y.; et al. NTIRE 2022 Spectral Recovery Challenge and Data Set. In Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), New Orleans, LA, USA, 19–20 June 2022; pp. 862–880. [CrossRef]
6. Liu, L.; Lei, S.; Shi, Z.; Zhang, N.; Zhu, X. Hyperspectral remote sensing imagery generation from RGB images based on joint discrimination. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2021**, *14*, 7624–7636. [CrossRef]
7. Du, D.; Gu, Y.; Liu, T.; Li, X. Spectral reconstruction from satellite multispectral imagery using convolution and transformer joint network. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 1–15. [CrossRef]
8. Deng, L.; Sun, J.; Chen, Y.; Lu, H.; Duan, F.; Zhu, L.; Fan, T. M2H-Net: A reconstruction method for hyperspectral remotely sensed imagery. *ISPRS J. Photogramm. Remote Sens.* **2021**, *173*, 323–348. [CrossRef]
9. Alibani, M.; Pastorino, M.; Moser, G.; Acito, N. Investigating the Potential of Deep Learning Approaches in the Reconstruction of VNIR-SWIR Hyperspectral Data From Multispectral Imagery. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2025**, *18*, 14215–14227. [CrossRef]
10. Heng, Y.; Wu, Y.; Chen, J.; Dasmahapatra, S.; Kim, H. MatSpectNet: Material Segmentation Network with Domain-Aware and Physically-Constrained Hyperspectral Reconstruction. *arXiv* **2023**, arXiv:2307.11466.
11. Gao, L.; Hong, D.; Yao, J.; Zhang, B.; Gamba, P.; Chanussot, J. Spectral Superresolution of Multispectral Imagery with Joint Sparse and Low-Rank Learning. *IEEE Trans. Geosci. Remote Sens.* **2021**, *59*, 2269–2280. [CrossRef]
12. Nasrabadi, N.M. Hyperspectral Target Detection : An Overview of Current and Future Challenges. *IEEE Signal Process. Mag.* **2014**, *31*, 34–44. [CrossRef]
13. Sneha.; Kaul, A. Hyperspectral imaging and target detection algorithms: A review. *Multimed. Tools Appl.* **2022**, *81*, 44141–44206. [CrossRef]
14. Wang, S.; Xu, Y.; Zeng, D.; Huang, F.; Liang, L. Deep learning-based spectral reconstruction in camouflaged target detection. *Int. J. Appl. Earth Obs. Geoinf.* **2024**, *126*, 103645. [CrossRef]
15. Cai, Y.; Lin, J.; Lin, Z.; Wang, H.; Zhang, Y.; Pfister, H.; Timofte, R.; Van Gool, L. MST++: Multi-Stage Spectral-Wise Transformer for Efficient Spectral Reconstruction. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, New Orleans, LA, USA, 19–24 June 2022; pp. 745–755.
16. Zamir, S.W.; Arora, A.; Khan, S.; Hayat, M.; Khan, F.S.; Yang, M.H.; Shao, L. Learning enriched features for fast image restoration and enhancement. *IEEE Trans. Pattern Anal. Mach. Intell.* **2022**, *45*, 1934–1948. [CrossRef] [PubMed]

17. Chen, B.; Liu, L.; Zou, Z.; Shi, Z. Target Detection in Hyperspectral Remote Sensing Image: Current Status and Challenges. *Remote Sens.* **2023**, *15*, 3223. [CrossRef]
18. Manolakis, D.; Pieper, M.; Truslow, E.; Cooley, T.; Brueggeman, M.; Lipson, S. The remarkable success of adaptive cosine estimator in hyperspectral target detection. In *Proceedings of the Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery XIX*; Shen, S.S.; Lewis, P.E., Eds.; International Society for Optics and Photonics, SPIE: Bellingham, WA, USA, 2013; Volume 8743, p. 874302. [CrossRef]
19. Chang, C.I. *Hyperspectral Imaging: Techniques for Spectral Detection and Classification*; Springer Science & Business Media: Berlin/Heidelberg, Germany, 2003; Volume 1.
20. Laboratory, J.P. MAIA Multispectral Camera. Available online: https://www.salengineering.it/public/en/p/maia.asp?utm_source (accessed on 12 November 2025).
21. Haneishi, H.; Hasegawa, T.; Hosoi, A.; Yokoyama, Y.; Tsumura, N.; Miyake, Y. System design for accurately estimating the spectral reflectance of art paintings. *Appl. Opt.* **2000**, *39*, 6621–6632. [CrossRef] [PubMed]
22. Imai, F.H.; Berns, R.S. Spectral estimation using trichromatic digital cameras. In *Proceedings of the International Symposium on Multispectral Imaging and Color Reproduction for Digital Archives*; Society of Multispectral Imaging of Japan: Chiba, Japan, 1999; Volume 42, pp. 1–8.
23. Cheung, V.; Westland, S.; Li, C.; Hardeberg, J.; Connah, D. Characterization of trichromatic color cameras by using a new multispectral imaging technique. *J. Opt. Soc. Am. A* **2005**, *22*, 1231–1240. [CrossRef] [PubMed]
24. Arad, B.; Ben-Shahar, O. Sparse recovery of hyperspectral signal from natural RGB images. In *Proceedings of the Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, 11–14 October 2016, Proceedings, Part VII 14*; Springer Berlin/Heidelberg, Germany, 2016; pp. 19–34.
25. Li, T.; Gu, Y. Progressive spatial–spectral joint network for hyperspectral image reconstruction. *IEEE Trans. Geosci. Remote Sens.* **2021**, *60*, 1–14. [CrossRef]
26. Cai, Y. MST-plus-plus. Available online: https://github.com/caiyuanhao1998/MST-plus-plus/tree/master (accessed on 17 November 2025).
27. Strang, G.; et al. *Linear Algebra and Learning from Data*; Wellesley-Cambridge Press: Wellesley, MA, USA, 2019; Volume 4.
28. Laboratory, J.P. AVIRIS-NG Data Portal. 2024. Available online: https://avirisng.jpl.nasa.gov/dataportal/ (accessed on 11 June 2025).
29. Ientilucci, E.J.; Adler-Golden, S. Atmospheric compensation of hyperspectral data: An overview and review of in-scene and physics-based approaches. *IEEE Geosci. Remote Sens. Mag.* **2019**, *7*, 31–50. [CrossRef]
30. ESA. 2025. Available online: https://sentinels.copernicus.eu/-/copernicus-sentinel-2c-spectral-response-functions (accessed on 12 June 2025).
31. Yokoya, N.; Grohnfeldt, C.; Chanussot, J. Hyperspectral and multispectral data fusion: A comparative review of the recent literature. *IEEE Geosci. Remote Sens. Mag.* **2017**, *5*, 29–56. [CrossRef]
32. Chang, C.I. Constrained Energy Minimization (CEM) for Hyperspectral Target Detection: Theory and Generalizations. *IEEE Trans. Geosci. Remote Sens.* **2024**, *62*, 1–21. [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.