Spectral Reconstruction for Enhanced Target Detection in Multispectral Imagery

1st Michael Alibani
DII, Uninversity of Pisa
56122 Pisa, Italy
michael.alibani@unipi.it

2nd Alessandro Terracciano
Italian Naval Academy
57127 Livorno, Italy
a.terracciano@studenti.unipi.it

3th Marco Diani
Italian Naval Academy
57127 Livorno, Italy
m.diani@iet.unipi.it

4rd Nicola Acito
DII, Uninversity of Pisa
56122 Pisa, Italy
nicola.acito@unipi.it

***Abstract*—Hyperspectral imaging offers high spectral resolution for accurate material discrimination and target detection, yet it is hindered by low spatial resolution and high operational costs. In this work, we present a novel framework that leverages deep learning–based spectral reconstruction to convert multispectral imagery into hyperspectral data with enhanced spectral and spatial resolution, specifically tailored for target detection. Our approach considers a spectral reconstruction network named MST++ that has been initially trained on multispectral–hyperspectral image pairs generated from AVIRIS-NG by means of proper Spectral Response Functions. A targeted fine-tuning procedure is then applied to optimize the recovery of discriminative spectral features for small, camouflaged targets. For the detection stage, reference-based matched filters exploit the prior spectral signatures of the targets. Validation on real multispectral data acquired by a UAV in field conditions demonstrates that our method significantly enhances the detection of small targets, achieving performance metrics intermediate between conventional multispectral and true hyperspectral imagery. These results underscore the potential of spectral reconstruction techniques as a cost-effective solution to augment target detection capabilities in scenarios where high-end hyperspectral sensors are not available.**

***Index Terms*—Target Detection, Spectral Reconstruction, Hyperspectral, Multispectral.**

I. Introduction

Hyperspectral (HS) and multispectral (MS) imaging methodologies are distinguished by a range of distinctive properties, which result in a variety of advantages and limitations. HS sensors acquire data in a number of narrow, contiguous spectral bands, thereby providing detailed spectral information that is indispensable for accurate material discrimination and target identification [1],[2],[3]. However, this high spectral resolution is typically associated with lower spatial resolution and significantly higher sensor and operational costs. Conversely, MS sensors capture data in a limited number of broader spectral bands, offering superior spatial resolution and reduced cost, but at the expense of spectral discriminability. This trade-off between resolutions (spatial and spectral) and cost is what motivates the pursuit of methodologies capable of synthesizing the spectral fidelity of HS imagery and the spatial precision and cost-efficiency of MS data. A promising approach is the application of spectral reconstruction (SR) techniques, which aim to generate HS-like data from inputs with limited number of bands [4]. Such methods provide a means to obtain imagery that exhibits both high spectral and high spatial resolution without necessitating the deployment of expensive HS sensors, which are often unavailable in many operational scenarios. Recent advances in deep learning have yielded several SR algorithms capable of reconstructing detailed hyperspectral information from lower-dimensional data sources, including both RGB [5],[6] and MS images [7],[8]. Despite the fact that this problem is inherently ill-posed, these techniques have been proven to be effective for downstream tasks such as image segmentation [9], classification [7] and unmixing [10]. Although the literature predominantly emphasizes the effectiveness of SR methods in segmentation and classification, relatively little attention has been paid to their impact on target detection, which is critical in both military and civilian applications [11]. Target detection, particularly of small objects, presents unique challenges because the subtle spectral differences that discern targets from complex backgrounds are often obscured in conventional MS images [12]. The recent work of Wang et al. [13] proposes an approach to reconstruct MS from RGB data aimed to improve the detection of extended (spatially) and camouflaged targets. In the present study, we extend the aforementioned research to detect small targets by leveraging SR to obtain high spectral resolution HS images from MS data in the Visible and Near-InfraRed (VNIR) spectral range. The approach is specifically designed to improve the performance of Spectral Signature-Based Target Detection (SSBTD) in MS images. It uses a state-of-the-art deep learning–based SR method, namely MST++ [14] followed by the well-known Normalized Matched Filter (NMF, [15],[16]). We propose a learning strategy based on simulated data for the SR model. It includes a specific Fine-Tuning (FT) procedure to optimize the reconstruction of the spectral signature of the target of interest. Preliminary experimental results are shown that have been obtained on an image acquired by a drone-based HS camera. Specifically, starting from the HS data, an MS image is obtained by considering the Spectral Response Functions (SRFs) of 9 Sentinel-2 bands in the VNIR spectral range. The MS image obtained serves as an input for the test of the proposed procedure. The real data adopted are equipped with a ground-truth map and spectral signatures of the targets considered.

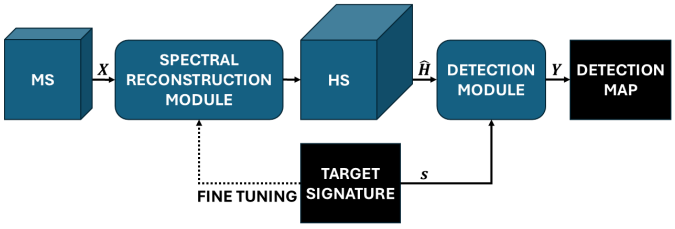
The work is organized as follows. Section II describes the proposed approach and the training paradigm. Section III describes the experiments performed and presents the results

obtained. Finally, Section IV summarizes the main conclusions drawn.

II. Proposed Method

A. Overview of the proposed approach

The processing chain of the proposed paradigm is summarized in Figure 1a. It basically includes two modules. The first, named the SR stage, implements a deep learning model based on attention and transformer mechanism to reconstruct the HS image from the MS input. The second, named the detection module, is an SSBTD algorithm whose detection statistic after thresholding provides the output detection map. A key aspect of our approach is the fine-tuning procedure represented by the dashed line in Figure 1a. It is specifically tailored to the target of interest and aims at specializing the SR model to the reconstruction of the spectral signature of that target. Fine-tuning is motivated by the fact that target materials are often rare and may not be well represented in the training set. The details of the SR module, the training procedures, and the fine-tuning approach are described in the following sections.



(a)

Fig. 1. Overview of the Proposed workflow. The diagram illustrates the complete process flow, starting with MS data input, followed by HS reconstruction, and ending in a detection module that generates the final decision map.

B. Hyperspectral reconstruction

Data-driven methods have rapidly gained traction due to their ability to capture intricate spectral patterns directly from data. For instance, several studies have demonstrated the effectiveness of convolutional neural networks and transformer-based architectures for spectral recovery task of solving the following problem:

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

where $\hat{\mathbf{H}} \in \mathbb{R}^{S \times W \times C}$ 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 the spectral dimension of the images (with $c \ll C$). The function $f(\cdot; \boldsymbol{\Theta})$, defined by the neural model's architecture, represents the parametric mapping whose parameters $\boldsymbol{\Theta}$ are learned during the training process by numerically solving the following optimization problem:

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

where L indicates loss function used for the optimization and $\mathbf{H} \in \mathbb{R}^{S \times W \times C}$ is the reference HS image.

For the purposes of this work, the SR module implements the MST++ network that resulted the winner at the NITRE Spectral Recovery challenge [5] for RGB to 31-band reconstruction.

MST++ employs a transformer-based architecture to capture long-range dependencies within the spectral domain. A key innovation of MST++ is the treatment of each spectral channel as an individual token. This design enables the network to perform fine-grained inter-channel comparisons using multi-head self-attention, thus effectively resolving subtle spectral details. Its multi-scale design allows for analyzing features at different resolutions, ensuring robustness against varying levels of data granularity. The loss function L adopted in the training phase of MST++ is the Mean Relative Absolute Error(MRAE) [5]. Of course, MST++ architecture was 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. This adaptation involved modifying the input and output interfaces, as well as resizing the dimensions of the multi-head attention layers to accommodate the increased spectral dimensionality.

C. Target detection

The target detection module processes the estimated HS image along with a known reference signature that represents the target of interest. For the purposes of this work the NMF algorithm is adopted. It is widely recognized for its robustness and optimality in maximizing the signal-to-noise ratio under Gaussian noise conditions [15], [16].

The NMF detection statistic for each pixel is computed as:

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

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

where $\| \cdot \|$ is the vector norm, $\mathbf{s} \in \mathbb{R}^C$ denotes the target spectral signature, $\mathbf{x}_i \in \mathbb{R}^C$ denotes the ith image pixel, $\mathbf{m}_b \in \mathbb{R}^C$ and $\mathbf{C}_b \in \mathbb{R}^{CxC}$ are the mean vector and the covariance matrix of the background, respectively. According to Eq. 3, the detection statistic y_i for the ith pixel measures the cosine of the angle between the target spectral signature and the pixel under test in the background whitened space. According to that geometrical interpretation y_i has values in the range $[-1, 1]$ where higher values correspond to a greater similarity between the target and the pixel under test. Both the background covariance matrix, $\mathbf{C}_b$, and the background mean vector, $\mathbf{m}_b$, are estimated from the entire image pixels using robust statistical techniques that minimize the impact of potential target pixels.

D. Training strategy

The training phase of the SR model requires a vast and diverse dataset, including both MS and HS images. However, the availability of such a dataset is highly unlikely in practical applications. Therefore, we propose obtaining the necessary data through simulation. Specifically, starting from a large repository of high spectral resolution HS images and assuming

the availability of the SRFs of both the MS sensor and the desired HS sensor, the MS-HS image pairs are generated by applying the aforementioned SRFs to each original high-resolution HS image. For this purpose, the freely accessible AVIRIS-NG image repository serves as a highly valuable resource. AVIRIS-NG is an airborne HS sensor developed by the Jet Propulsion Laboratory (JPL) that delivers high-resolution imagery both spatially and spectrally. AVIRIS-NG measures the incoming radiation at wavelengths ranging from 380 nm to 2510 nm at a spectral resolution of 5 nm. It captures images with 600 cross-track samples with spatial sampling from 0.3 m to 7.0 m, and it is ideal for simulating the image pairs required for our work.

From the MS-HS image pairs generated according to the aforementioned procedure, N image tiles of size 32×32 pixels are extracted to train the *general* SR model. Then, using a fine-tuning paradigm, specialized SR models are obtained for each target of interest. Detecting small targets often involves materials that are rarely represented in remote sensing images compared to abundant background features. Although the large dataset used for general-purpose reconstruction benefits overall model performance, it may not adequately represent these rare targets. Therefore, after the initial training, we propose to perform a fine-tuning procedure on a smaller dataset. It is obtained by randomly selecting $n \ll N$ pairs of tiles from the original training set, where the spectral signature of the target is artificially injected and then spectrally blended with the background. The target's position and size (in pixels) within each MS-HS pair are randomly chosen to cover up to 20% of the tile size. It is worth noting that fine-tuning is performed on a reduced number of training samples, and this makes the learning task less time-consuming than training the general-purpose SR model.

III. Experimental Analysis

A. Experimental dataset

For the experimental evaluation of our proposed approach in a real scenario, we assembled a comprehensive dataset collected during a dedicated acquisition campaign held in Italy and using a *Headwall Nano Hyperspec-VNIR* camera mounted on a UAV. The camera is a portable instrument that acquires 270 spectral samples within the 400–1000 nm spectral range (VNIR). Those spectral samples are generally binned with a factor of 3 to increase the Signal to Noise Ratio (SNR) thus obtain HS images with $C = 90$ bands. During the campaign we executed flights over a measurement field in which we deployed several targets across the terrain. The flights were performed at a 27 m altitude, with a ground sample distance (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 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 experiment presented in this work concerns one of the acquired image whose RGB representation is reported in Figure 2a. The locations of the considered targets in the image are highlighted in Figure 2a. Figures 2b–2d show pictures of the considered targets, which include a plastic yellow plane (Target 1), a plastic green tank (Target 2), a plastic fake plant (Target 3), and a colored mimetic wooden board (Target 4). In order to test the proposed procedure, starting from the 90-band HS image we simulated a 9-band MS image. To this purpose, we employed the SRFs of the channels in the VNIR spectral range of an existing MS sensor, i.e. Sentinel-2.

(a) Measurement Field

(b) Target 1 (c) Target 2 (d) Target 3 (e) Target 4

Fig. 2. Detail of the measurement field. (a) RGB view of the measurement field acquired from 27 m using the aerial platform. Detailed views of the targets: (b) target 1, yellow plastic plane; (c) target 2, green plastic tank; (d) target 3, fake plastic plant; and (e) target 4, mimetic-colored board.

B. Training and fine-tuning

To simulate the training set for the general-purpose SR model we selected images acquired by AVIRIS-NG sensors from various regions worldwide, including Europe, the USA, and Mexico and spanning the period 2018—2022. These datasets cover diverse scenarios (urban, rural, coastal, and mountainous areas). According to the strategy described in Section II-D, starting from the selected AVIRIS-NG data and using the SRFs of both the 9 Sentinel-2 bands and the HS sensor adopted in the measurement campaign, MS and HS image pairs were simulated. From the resulting dataset, we extracted $N = 6000$ tiles of size 32×32 pixels. Training was performed on an NVIDIA GeForce 4090 GPU (24GB RAM) using the Adam optimizer (with $\beta_1 = 0.9$ and $\beta_2 = 0.99$). The learning rate was initially set to $l_r = 10^{-4}$ and dynamically adjusted using a cosine annealing schedule. The network was trained for 200 epochs, with a total time of approximately 24 hours. For each of the considered targets we generated $n = 600$ MS-HS tiles pairs by injecting the corresponding spectral signature measured by the spectroradiometer. With reference to Target 1, Figure 3 shows and example of the tile used for the fine-tuning. In the figure the target is highlighted by the yellow box. The fine-tuning step was performed on the same GPU, with a lower learning rate ($l_r = 10^{-5}$) and for 10 epochs, which took only about 10 minutes.

C. Experimental results and discussion

This section presents the results of the enhanced MS target detection aided by the SR network. To quantitatively assess

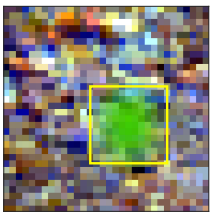


Fig. 3. RGB representation of a 32×32 patch used in the fine-tuning process. The yellow box highlights the injected fake target, which has been spectrally mixed with the background.

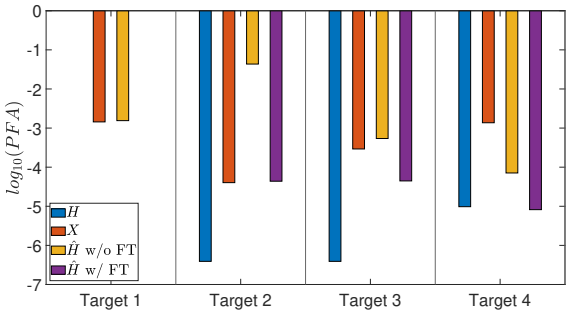


Fig. 4. Comparison of detection results on a logarithmic scale for four targets, indicating the probability of false alarm when 50% of the targets are detected. Blue bars represent detection using ground truth HS data; red bars show results for MS data; yellow bars indicate detection on reconstructed HS data; and purple bars represent detection on reconstructed HS data with fine-tuning. The lower the value, the better the result. Note: an empty bar indicates no false alarms (best result: PFA=0).

the performance of our proposed method, we first evaluate the capability of the general-purpose SR network to spectrally reconstruct the HS image using the Relative Root Mean Square Error (RRMSE) evaluation metric, which can be interpreted as the pixel-wise percentage error in the reconstruction of all spectral bands relative to the corresponding reference pixels. RRMSE is defined as

$$\text{RRMSE}(\mathbf{H}, \hat{\mathbf{H}}) = \frac{\sqrt{\sum_{i=1}^{K}(H_i - \hat{H}_i)^2}}{\sqrt{\sum_{i=1}^{K} H_i^2}} \tag{4}$$

where $K = S \times W \times C$ indicates the cardinality of the data cube while H_i and $\hat{H}_i$ are the ith pixels in $\mathbf{H}$ and $\hat{\mathbf{H}}$, respectively.

Eq. 4 shows that RRMSE is evaluated on the HS image obtained by the SR module taking as reference the original HS image. We obtain a mean RRMSE = 0.032 for the overall image, meaning that the reconstruction error is in average in the order of 3.2%. It should be noted that the network commits higher error in regions corresponding to tree shadows, which are associated with a lower SNR.

Secondly we evaluate the target detection performance in terms of Probability of False Alarm (PFA) that measures the likelihood that the detection system will indicate a detection (a positive) when there is actually no target present. Specifically, for each target, we computed the threshold value corresponding to the detection of 50% of the pixels occupied by that target in the image. Then, we calculated the PFA as the fraction of all image pixels for which the NMF statistic exceeds the derived threshold.

Figure 4 shows the results in terms of PFA (in logarithmic scale) for the four considered targets. Specifically, Figure 4, for each target, shows $\log_{10}(\text{PFA})$ by applying the NMF algorithm to a) the original 90-band HS image (blue bar), b) the 9-band MS image (red bar), c) the HS image reconstructed by the general-purpose SR network (yellow bar), and HS image after fine-tuning (purple bar). The expectation is that detection on the reconstructed target with specific fine-tuning provides a better (i.e., lower) PFA compared to detection on the MS image, but a worse value than detection on the reference HS. In all cases, the $\log_{10}(\text{PFA})$ using the MS (red bars) is higher than that obtained with the ground truth (blue bars). Notice that an empty bar indicates that there were no false alarms ($\log_{10}(\text{PFA}) = -\infty$). Detection on the SR HS data (yellow bars) yields at least a detection rate similar to the MS case (except for Target 2), while detection on the SR HS data with fine-tuning improves performance in all cases except for Target 2, where it remains equal to the MS case. Notably, for Targets 1 and 4, the detection on the fine-tuned data achieves performance equivalent to that of the original HS data. This result highlights the capability of this approach to improve target detection on MS data by leveraging the power of deep spectral reconstruction networks, and it encourages the research in the direction presented in this work.

IV. Conclusions

In this paper, we introduced a novel operational workflow to enhance small target detection in multispectral data by leveraging advanced spectral reconstruction and target detection techniques. Specifically, we employed the MST++ method to reconstruct hyperspectral images from multispectral ones, and we applied a Normalized Matched Filter for target detection. An *ad hoc* fine-tuning procedure was also integrated to optimize the deep learning-based spectral reconstruction process, particularly for the detection of rare target materials.

Experimental validation on drone-acquired field data demonstrated that our approach significantly improves detection performance, particularly when assessed using the probability of false alarm when 50% of the targets are detected as the key metric. In most cases, detection on the reconstructed hyperspectral images outperformed that on the original multispectral data, underscoring the effectiveness of our integrated workflow.

These promising results pave the way for further research. Future work will focus on extending the approach to a broader range of targets and operational conditions, with the aim of further refining the methodology and validating its applicability in diverse real-world scenarios.

ACKNOWLEDGMENT

This work was supported in part by the University of Pisa under Grant PRA 2022 91 INTERCONNECT

REFERENCES

[1] D. G. Manolakis, R. B. Lockwood, and T. W. Cooley, *Hyperspectral imaging remote sensing: physics, sensors, and algorithms*. Cambridge University Press, 2016.

[2] E. Bedini, "The use of hyperspectral remote sensing for mineral exploration: A review," *Journal of Hyperspectral Remote Sensing*, vol. 7, no. 4, pp. 189–211, 2017.

[3] N. Acito, M. Diani, M. Alibani, and G. Corsini, "Matched filter based on the radiative transfer model for co2 estimation from prisma hyperspectral data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 61, pp. 1–13, 2023. DOI: 10.1109/TGRS.2023.3327756.

[4] J. Zhang, R. Su, Q. Fu, W. Ren, F. Heide, and Y. Nie, "A survey on computational spectral reconstruction methods from rgb to hyperspectral imaging," *Scientific reports*, vol. 12, no. 1, p. 11 905, 2022.

[5] B. Arad, R. Timofte, R. Yahel, *et al.*, "Ntire 2022 spectral recovery challenge and data set," in *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2022, pp. 862–880. DOI: 10.1109/CVPRW56347.2022.00102.

[6] L. Liu, S. Lei, Z. Shi, N. Zhang, and X. Zhu, "Hyperspectral remote sensing imagery generation from rgb images based on joint discrimination," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 14, pp. 7624–7636, 2021.

[7] D. Du, Y. Gu, T. Liu, and X. Li, "Spectral reconstruction from satellite multispectral imagery using convolution and transformer joint network," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 61, pp. 1–15, 2023.

[8] L. Deng, J. Sun, Y. Chen, *et al.*, "M2h-net: A reconstruction method for hyperspectral remotely sensed imagery," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 173, pp. 323–348, 2021.

[9] Y. Heng, Y. Wu, J. Chen, S. Dasmahapatra, and H. Kim, *Matspectnet: Material segmentation network with domain-aware and physically-constrained hyperspectral reconstruction*, 2023. arXiv: 2307.11466 [cs.CV]. [Online]. Available: https://arxiv.org/abs/2307.11466.

[10] L. Gao, D. Hong, J. Yao, B. Zhang, P. Gamba, and J. Chanussot, "Spectral superresolution of multispectral imagery with joint sparse and low-rank learning," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 59, no. 3, pp. 2269–2280, 2021. DOI: 10.1109/TGRS.2020.3000684.

[11] N. M. Nasrabadi, "Hyperspectral target detection : An overview of current and future challenges," *IEEE Signal Processing Magazine*, vol. 31, no. 1, pp. 34–44, 2014. DOI: 10.1109/MSP.2013.2278992.

[12] N. Yokoya, C. Grohnfeldt, and J. Chanussot, "Hyperspectral and multispectral data fusion: A comparative review of the recent literature," *IEEE Geoscience and Remote Sensing Magazine*, vol. 5, no. 2, pp. 29–56, 2017.

[13] S. Wang, Y. Xu, D. Zeng, F. Huang, and L. Liang, "Deep learning-based spectral reconstruction in camouflaged target detection," *International Journal of Applied Earth Observation and Geoinformation*, vol. 126, p. 103 645, 2024.

[14] Y. Cai, J. Lin, Z. Lin, *et al.*, "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*, Jun. 2022, pp. 745–755.

[15] D. Manolakis and G. Shaw, "Detection algorithms for hyperspectral imaging applications," *IEEE Signal Processing Magazine*, vol. 19, no. 1, pp. 29–43, 2002. DOI: 10.1109/79.974724.

[16] C.-I. Chang, *Hyperspectral imaging: techniques for spectral detection and classification*. Springer Science & Business Media, 2003, vol. 1.