Guided Deep Generative Model-Based Spatial Regularization for Multiband Imaging Inverse Problems

Min Zhao, Nicolas Dobigeon, Jie Chen

Guided Deep Generative Model-based Spatial Regularization for Multiband Imaging Inverse Problems

Min Zhao, *Student Member, IEEE,* Nicolas Dobigeon, *Senior Member, IEEE* and Jie Chen, *Senior Member, IEEE*

***Abstract*—When adopting a model-based formulation, solving inverse problems encountered in multiband imaging requires to define spatial and spectral regularizations. In most of the works of the literature, spectral information is extracted from the observations directly to derive data-driven spectral priors. Conversely, the choice of the spatial regularization often boils down to the use of conventional penalizations (e.g., total variation) promoting expected features of the reconstructed image (e.g., piece-wise constant). In this work, we propose a generic framework able to capitalize on an auxiliary acquisition of high spatial resolution to derive tailored data-driven spatial regularizations. This approach leverages on the ability of deep learning to extract high level features. More precisely, the regularization is conceived as a deep generative network able to encode spatial semantic features contained in this auxiliary image of high spatial resolution. To illustrate the versatility of this approach, it is instantiated to conduct two particular tasks, namely multiband image fusion and multiband image inpainting. Experimental results obtained on these two tasks demonstrate the benefit of this class of informed regularizations when compared to more conventional ones.**

***Index Terms*—Multiband imaging, inverse problems, deep learning, deep image prior, guided image, deep generative regularization.**

I. Introduction

DUE to an unavoidable yet inherent spatial vs. spectrum trade-off, multiband images are generally characterized by a low spatial resolution, which is far lower than that of more conventional imaging techniques, such as red-green-blue (RGB) images [1]. Moreover, depending on the considered applicative contexts, the observed images may be corrupted by a high level noise or affected by undersampling or blurs. Part of the measurements may also be unavailable when the acquisition of the full data cube is impossible. This can arise in several applicative scenarios due to constraints imposed by the instrumental process or the measurement protocol. For example, during image acquisition by scanning transmission electron microscopy (STEM), the electron beam can induce sample damage for sensitive materials. To preserve the imaged sample, random sampling schemes are implemented, i.e., only a small part of the pixel locations is visited by the probe, which results in a reduction of the total acquisition time and of the energy received by the sample [2]. However, this partial acquisition procedure prevents any subsequent analysis, such as target detection, material identification, and classification.

To overcome these limitations, one popular approach consists in restoring a multiband image of full or higher spatial resolution from the degraded measurements. This task can be formulated as solving an inverse problem that aims at recovering a full image of higher quality. Denoising, deconvolution, inpainting, single-image super-resolution and multiple-image fusion are some archetypical examples of such inversion tasks dedicated specifically to multiband imaging.

Numerous works were devoted to the design of multiband imaging inversion techniques. They can be mainly categorized into conventional model-based and more recent learning-based methods. The former address the problem by minimizing a data fitting term coupled with a handcrafted regularization to promote predefined characteristics of the restored image. This strategy has shown to be particularly appealing to capture the spectral redundancy of the images, e.g., by imposing a low rank structure. It is worth noting that this structure can be informed by analyzing the spectral content of the acquired multiband image itself, e.g., by conducting a principal component analysis (PCA) [3], [4]. Regarding the spatial regularizations, numerous handcrafted model-based priors have been proposed, such as total variation [5], sparsity [6], low-rankness [7] and dictionary learning [3]. However, selecting an appropriate regularizer to match the intrinsic properties of the image is a nontrivial task. More importantly, these models can hardly incorporate the richness of the spatial content of the images. Devising sophisticated but sufficiently generic spatial regularizations able to capture the diversity of the image properties is generally accompanied by a significant increase of the resulting computational burden.

Conversely learning-based methods have been recently proposed to circumvent this bottleneck and have became a hotspot thanks to its superior capability to excavate high-level features. These approaches aim at learning a nonlinear mapping from the raw measurements to the restored image where the image priors are implicitly encoded in the network parameters. However, learning-based methods generally require large training data sets. This may impair their use for non-conventional imaging techniques for which such data sets may be of limited availability. Moreover, such black-box techniques generally lacks of physical interpretability.

Recently, combining conventional model-based and learning-based methods has shown to be a promising way to

M. Zhao and J. Chen are with School of Marine Science and Technology, Northwestern Polytechnical University, Xi'an 710072, China (e-mail: minzhao@mail.nwpu.edu.cn; dr.jie.chen@ieee.org).

Nicolas Dobigeon is with University of Toulouse, IRIT/INPENSEEIHT, CNRS, 2 rue Charles Camichel, BP 7122, 31071 Toulouse Cedex 7, France and also with the Institut Universitaire de France (IUF), France (e-mail: Nicolas.Dobigeon@enseeiht.fr).

overcome the respective limitations inherent of each approach. Following this trend, this paper proposes a novel and smart framework to spatially regularize multiband imaging inverse problems. Particularly, the proposed method incorporates a learning-based spatial regularization into a conventional model-based formulation. It thus leverages the power of deep learning able to encode complex image characteristics and the advantages of a physical-based model exploiting the physical measurement process. More precisely, the proposed regularization consists of a pretrained deep generative decoder informed by an auxiliary image of high spatial resolution, which can be interpreted as a domain-adapted prior. Indeed this paper focuses on appealing scenarios where the imaging protocol provides a higher-resolution image of the same observed scene to train the deep regularization. Such scenarios arise naturally in various applicative contexts. This is the case when images of complementary spatial and spectral resolutions have to be fused or when the acquisition of a multiband image to be restored is concomitant with the acquisition of an auxiliary image of higher spatial resolution. The main contributions of this work can be summarized as follows:

- We propose a new way to regularize multiband imaging inverse problems by means of a deep generative network able to encode prior information learnt from a high spatial resolution complementary image. As an example, this generative network is chosen as a guided deep decoder (GDD) recently proposed in the literature. However, the proposed framework may not be limited to this particular choice and may remain valid for any guided generative model that would be designed in future works. The relevance of this informed regularization is illustrated by comparing experimental results with those obtained when the generative model is chosen as a variational encoder (VAE) not trained on the auxiliary image specifically.
- Instead of resorting to an automatic differentiation technique to minimize the resulting optimization criterion, we devise a splitting-based strategy which has the great advantages of decomposing the initial problem into simpler subproblems. In particular for some subproblems, closed-form algorithmic updates can be implemented, which significantly reduce the computational time. We empirically validate this choice by comparing the restoration results reached by the proposed strategy with those obtained by the Adam optimizer.
- To illustrate the versatility of the proposed framework, it is instantiated for two ubiquitous inversion tasks, namely image fusion and image inpainting. Extensive simulation results obtained from experiments conducted for these two tasks show that the proposed framework competes favorably with respect to state-of-the-art inversion methods.

The remainder of this paper is organized as follows. Recent works about multiband imaging inversion are reviewed in Section II, with a particular focus on the design of model-based and learning-based regularizations. The problem addressed in this paper is stated in Section III. Section IV introduces the proposed generic framework to perform multiband imaging inversion. This framework is instantiated in Section V for two particular yet ubiquitous tasks, namely fusion and inpainting. Section VI reports extensive experimental results obtained for these two tasks. Section VII concludes the paper.

II. Related Works

A. Model-based regularizations

A significant amount of regularizations have been designed to describe the underlying image characteristics and improve the inversion task. Reviewing the whole literature would be a titanic task and is out-of-the-scope of this paper. Only a few directions are listed in the sequel of this section. Numerous works such as [5], [8] and [9] use a conventional total variation (TV) to preserve edges and promote piece-wise constant content. The methods [3], [6] impose low-rank structures of the multiband image, e.g., by decomposing the data into a set of basis elements or dictionary. The representation coefficients are then associated to sparsity-promoting penalizations, such as the ℓ_0-pseudo-norm or the ℓ_1-norm. The intrinsic property of image (non-local) self-similarity is another key strategy to exploit the redundancy arising in multiband images. The work [6] proposes a sparse low-rank representation model to perform multiband image super-resolution. In [10], a 3D non-local sparse representation is introduced to take advantage of non-local similarity in both spatial and spectral domains. Another strategy consists in resorting to a superpixel segmentation step to group spectrally similar pixels. The work [11] exploits the entropy rate superpixel segmentation method to divide the image into superpixels that are subsequently processed to ensure spectral smoothness and preserve image details.

B. End-to-end learning-based methods

To avoid designing handcrafted model-based priors, one alternative consists in formulating the inversion problem as a learning task by leveraging on the generalization ability of deep neural networks. It is then expected that the spatial and spectral redundancies intrinsic to the images are learnt from the training data set to be subsequently used as an implicit prior while solving the inversion problem. In [12], the authors design a deep convolutional neural network (CNN) for hyperspectral image restoration, which uses a modified U-net and decomposes the 3D filtering into 2D spatial filtering and 1D spectral filtering to reduce the number of parameters. In [13], a channel attention mechanism is used to capture spectral correlation information, and a local discriminative network is proposed to exploit a certain spatial continuity. The authors in [14] introduce a generative adversarial network to perform a pan-sharpening task. In [15], a self-supervised algorithm is proposed for hyperspectral image restoration. A 2D image denoiser pretrained on gray or RGB images is used as a backbone model and then fine-tuned on hyperspectral image band-by-band. The work [16] proposes an unsupervised deep framework for hyperspectral super-resolution. In [17], a deep hyperspectral prior algorithm is designed for hyperspectral restoration. It is based on 3D convolutional networks able to jointly learn the spatial and local spectral information.

C. Embedding learnt image prior into model-based methods

Deep architectures are known to demonstrate a certain superiority in extracting image properties efficiently. However end-to-end deep learning methods generally lack from explicability and, more importantly, do not capitalize on the knowledge about the acquisition process. Besides using handcrafted priors, a new trend consists in embedding a prior knowledge learnt by the deep networks into more conventional model-based iterative optimization algorithm. The works [18]–[20] exploit the output of deep networks as a constrained term to regularize the optimization problem. This regularizer term is chosen simple and convex, such as the squared Euclidean distance between the trained solution and the target estimation, which avoids heavy computation. Plug-and-play priors have also received a great attention in the context of multiband imaging inversion. The state-of-the-art denoising algorithms, such as CNN-based denoisers, are usually plugged into this framework as the proximity operator to capture the instinct spatial structures of images. For instance, the works [4], [21] decompose the optimization problem into iterative subproblems. Specifically, one of the subproblem can be cast as a proximal mapping related to the image prior model. Based on this interpretation, this subproblem can then be solved using a deep denoising operator, which incorporates deep priors into the estimation. Unrolling state-of-the-art optimization-based algorithms is another route followed by several recent works. It unfolds iterative optimization algorithms to derive a counterpart on the form of a trainable deep architectures. It allows the involved parameters to be learnt with the restored image jointly. For example, the work [22] unrolls alternating direction methods of multipliers (ADMM) as deep networks to perform hyperspectral super-resolution, and the work [23] unfolds the proximal gradient algorithm into deep networks for multiband image fusion.

III. Problem statement

This work considers a set $\mathbf{Y} = \{\mathbf{Y}_1, \ldots, \mathbf{Y}_K\}$ of K acquired multiband images $\mathbf{Y}_k \in \mathbb{R}^{B_k \times N_k}$ $(k = 1, \ldots, K)$ where $B_k \geq 1$ and $N_k \geq 1$ denote the numbers of bands (or channels) and pixels, respectively. These observations are assumed to be related to an unknown (latent) image $\mathbf{X} \in \mathbb{R}^{B \times N}$ through the direct model

$$\mathcal{H}(\mathbf{Y}) = \mathcal{M}(\mathbf{X}) + \mathbf{N} \tag{1}$$

where $\mathcal{M}(\cdot)$ represents the forward operators mapping from the latent space to the observation space. In this work, the operator underlying $\mathcal{M}(\cdot)$ is assumed to be perfectly known and can describe various spatial or spectral degradations including spatial blurring, regular or irregular spatial subsampling, spectral filtering, etc. The operator $\mathcal{H}(\cdot)$ aims at selecting and rearranging the data from the set $\mathbf{Y}$ to form the observations as provided by the sensors. For instance, it may select one multiband image from the K acquired images $\mathbf{Y}_1, \ldots, \mathbf{Y}_K$ when this unique image is intended to be a spatially and/or spectrally degraded version of the latent image $\mathbf{X}$. In (1), the matrix $\mathbf{N}$ stands for measurement noise and any mismodeling.

Remark (Complementary acquisitions) *In most cases and particularly in the applications considered in this paper (see Section V), the number of acquisitions is limited to $K = 2$ and the acquired images are of complementary spatial and spectral resolutions. One of these two images corresponds to a low spatial and high spectral resolution image and, to be more explicit, it will be denoted as $\mathbf{Y}_{\mathrm{HS}}$, while the other, denoted as $\mathbf{Y}_{\mathrm{HR}}$ is of high spatial and low spectral resolution, with $B_{\mathrm{HS}} \geq B_{\mathrm{HR}}$ and $N_{\mathrm{HS}} \leq N_{\mathrm{HR}}$.*

This paper addresses the problem of recovering the latent image $\mathbf{X}$, which is generally of the highest spatial and spectral resolutions, i.e., $N \geq \max_k\{N_k\}$ and $B \geq \max_k\{B_k\}$. This can be formulated as the optimization problem

$$\min_{\mathbf{X}} \|(\mathcal{H}(\mathbf{Y}) - \mathcal{M}(\mathbf{X}))\|^2_{\mathbf{\Sigma}^{-1}} + \mathcal{R}(\mathbf{X}) \tag{2}$$

where $\mathcal{R}(\cdot)$ is a regularization and the Mahalanobis norm $\|\cdot\|_{\mathbf{\Sigma}^{-1}}$ may account for particular measurement noise characteristics encoded in the covariance matrix $\mathbf{\Sigma}^{-1}$. This penalization function is often designed to be separable with respect to the spatial and the spectral information, i.e.,

$$\mathcal{R}(\mathbf{X}) = \mathcal{R}_{\mathrm{spa}}(\mathbf{X}) + \mathcal{R}_{\mathrm{spe}}(\mathbf{X}) \tag{3}$$

where the two terms on the right-hand side encode the expected spatial and spectral properties of $\mathbf{X}$, respectively. In most of the works dedicated to the restoration of multiband images, the pixels $\mathbf{x}_n \in \mathbb{R}^B$ $(n = 1, \ldots, N)$ of the unknown image $\mathbf{X}$ are assumed to live in a subspace $\mathbb{V} \subset \mathbb{R}^{\tilde{B}}$ of significantly lower dimension than the original space, i.e. $\tilde{B} \ll B$. This property can be promoted by choosing a spectral regularization $\mathcal{R}_{\mathrm{spe}}(\cdot)$ enforcing a low-rank structure on $\mathbf{X}$ by penalizing the rank

$$\mathcal{R}_{\mathrm{spe}}(\mathbf{X}) = \lambda_{\mathrm{spe}}\ \mathrm{rank}\,(\mathbf{X}) \tag{4}$$

or its convex relaxation, i.e., the nuclear norm

$$\mathcal{R}_{\mathrm{spe}}(\mathbf{X}) = \lambda_{\mathrm{spe}}\ \|\mathbf{X}\|_* \tag{5}$$

where λ_{spe} is a hyperparameter adjusting the weight of the regularization. One data-driven alternative consists in estimating the signal subspace $\mathbb{V}$ and its dimension $\tilde{B}$ beforehand from the image of highest spectral resolution $\mathbf{G}_{\mathrm{HS}}$ available in the set $\mathbf{Y}$, i.e., $\mathbf{G}_{\mathrm{HS}} \in \mathbf{Y}$. This subspace estimation is generally conducted by a principal component analysis [3], [4] or by using a dedicated subspace identification strategy [24], [25]. Then the spectral regularization could be defined as

$$\mathcal{R}_{\mathrm{spe}}(\mathbf{X}) = \sum_{n=1}^{N} \iota_{\mathbb{V}}(\mathbf{x}_n) \tag{6}$$

where $\iota_{\mathbb{V}}(\cdot)$ is the indicator function on the set $\mathbb{V}$. However, to simultaneously reduce the computational complexity of the resulting algorithms, a widely admitted strategy consists in imposing the factorization

$$\mathbf{X} = \mathbf{V}\mathbf{A} \tag{7}$$

where $\mathbf{V} \in \mathbb{R}^{B \times \tilde{B}}$ is a matrix whose $\tilde{B} \ll B$ columns span the lower dimensional subspace $\mathbb{V}$ and is generally chosen as orthonormal, i.e., $\mathbf{V}^\top\mathbf{V} = \mathbf{I}_{\widetilde{B}}$ where $\mathbf{I}_{\widetilde{B}}$ is the $\widetilde{B} \times \widetilde{B}$

identity matrix. The matrix $\mathbf{A} \in \mathbb{R}^{\tilde{B}\times N}$ contains the unknown representation coefficients of the pixels projected onto the subspace. Under this constraint, the original formulation (2) can be rewritten as an optimization problem with respect to the representation coefficients $\mathbf{A}$

$$\hat{\mathbf{A}} = \arg\min_{\mathbf{A}} \|\mathcal{H}(\mathbf{Y}) - \mathcal{M}(\mathbf{V}\mathbf{A})\|^2_{\mathbf{\Sigma}^{-1}} + \mathcal{R}_{\text{spa}}(\mathbf{A}) \quad (8)$$

with $\hat{\mathbf{X}} = \mathbf{V}\hat{\mathbf{A}}$. This latest formulation adopted by plenty of research works from the literature relies on an explicit data-driven spectral regularization specifically learnt from the observed image $\mathbf{G}_{\text{HS}}$ of highest spectral resolution which acts as a spectral guidance image. Conversely, very few attempts have been dedicated to the design of a data-driven spatial regularization $\mathcal{R}_{\text{spa}}(\cdot)$ exploiting the observed image of highest spatial resolution among the set of observations $\mathbf{Y}$. This paper aims at filling this gap by proposing a generic framework able to encode relevant spatial information into a deep generative model acting as a regularizer. This framework is described in the next section.

IV. Proposed Framework

This section describes the general framework specifically proposed to spatially regularize multiband image inverse problems when a high spatial resolution image is available in the set $\mathbf{Y}$. This image, denoted $\mathbf{G}_{\text{HR}} \in \mathbf{Y}$, acts as spatial guidance image for the data-driven spatial regularization. Its generic formulation and the corresponding algorithmic scheme are introduced in Sections IV-A and IV-B. This framework offers the possibility of embedding any existing deep generative model whose key feature is its ability to extract relevant spatial features from the spatial guidance image $\mathbf{G}_{\text{HR}} \in \mathbf{Y}$. It is worth noting that the choice of this network is left to the end-user who could select the most appropriate and up-to-date from the latest literature. In what follows, this framework is instantiated for one particular network as an illustrative purpose. Its architecture and the training strategy are detailed in Section IV-C.

A. Generic Formulation

Inspired by the so-called deep generative model [26] and deep image prior approach [27], the proposed framework leverages on the ability of deep networks to encode prior knowledge. More precisely, a generative decoder is trained to learn a mapping $\mathsf{D}_\theta(\cdot)$ from a latent space $\mathcal{Z}$ to the space $\mathbb{R}^{\tilde{B}\times N}$ of the representation coefficients $\mathbf{A}$, where θ represents the network parameters of the decoder. As the image $\mathbf{X}$ to be recovered is constrained to belong to the range $\mathbb{V}$ of the matrix $\mathbf{V}$ (see (7)), its representation coefficients $\mathbf{A}$ are assumed to belong to the range of the nonlinear mapping $\mathsf{D}_\theta(\cdot)$. This constraint can be satisfied by imposing

$$\mathbf{A} = \mathsf{D}_\theta(\mathbf{Z}) \quad (9)$$

where $\mathbf{Z} \in \mathbb{R}^{k\times N}$ is the latent representation matrix equipped with a Gaussian prior. Finally, the unknown image is estimated following

$$\hat{\mathbf{X}} = \mathbf{V}\mathsf{D}_\theta(\hat{\mathbf{Z}}) \quad (10)$$

where the estimated latent representation matrix $\hat{\mathbf{Z}}$ is the solution of the problem

$$\min_{\mathbf{Z}} \|\mathcal{H}(\mathbf{Y}) - \mathcal{M}(\mathbf{V}\mathsf{D}_\theta(\mathbf{Z}))\|^2_{\mathbf{\Sigma}^{-1}} + \lambda\|\mathbf{Z}\|^2_{\text{F}} \quad (11)$$

with λ a hyperparameter. The generic algorithmic scheme implemented to solve this problem is detailed in what follows.

B. Optimization

The optimization problem (11) can be challenging to solve, not only because of the nonlinearity induced by the mapping $\mathsf{D}_\theta(\cdot)$ but also because of the forward modeling $\mathcal{M}(\cdot)$. However it can be tackled efficiently by designing ADMM which allows the original problem to be decomposed into a 3-step procedure with simpler subproblems. By explicitly introducing the representation coefficient matrix $\mathbf{A}$, an equivalent constrained formulation writes

$$\min_{\mathbf{A},\mathbf{Z}} \|\mathcal{H}(\mathbf{Y}) - \mathcal{M}(\mathbf{V}\mathbf{A})\|^2_{\mathbf{\Sigma}^{-1}} + \lambda\|\mathbf{Z}\|^2_{\text{F}} \text{ s.t. } \mathbf{A} = \mathsf{D}_\theta(\mathbf{Z}). \quad (12)$$

Then the ADMM consists in iteratively performing the 3 following steps

$$\mathbf{A}^{(t+1)} = \arg\min_{\mathbf{A}} \|\mathcal{H}(\mathbf{Y}) - \mathcal{M}(\mathbf{V}\mathbf{A})\|^2_{\mathbf{\Sigma}^{-1}} + \mu \left\| \mathsf{D}_\theta(\mathbf{Z}^{(t)}) - \mathbf{A} + \frac{1}{2\mu}\mathbf{U}^{(t)} \right\|^2_{\text{F}} \quad (13)$$

$$\mathbf{Z}^{(t+1)} = \arg\min_{\mathbf{Z}} \left\| \mathsf{D}_\theta(\mathbf{Z}) - \mathbf{A} + \frac{1}{2\mu}\mathbf{U}^{(t)} \right\|^2_{\text{F}} + \frac{\lambda}{\mu}\|\mathbf{Z}\|^2_{\text{F}} \quad (14)$$

$$\mathbf{U}^{(t+1)} = \mathbf{U}^{(t)} + 2\mu\left(\mathsf{D}_\theta(\mathbf{Z}^{(t+1)}) - \mathbf{A}^{(t+1)}\right) \quad (15)$$

where $\mathbf{U}$ is a Lagrangian multiplier and μ is a penalty parameter. Interestingly, the minimization (13) stands for a generic formulation of an ℓ_2-regularized inverse problem. For most multiband imaging tasks, a closed-form solution can be implemented straightforwardly, as it will be shown for two ubiquitous tasks considered in Section V. The problem (14) is a nonlinear least-square problem similar to a projection onto the range of $\mathsf{D}_\theta(\cdot)$. In practice, it is empirically solved by resorting to a optimizer dedicated to deep learning, e.g., Adam. The overall algorithmic sketch of the proposed generic framework is summarized in Algorithm 1.

C. Guided deep decoder based generative model

As an illustrative instance, one particular network from the literature is considered to learn the generative model $\mathsf{D}_\theta(\cdot)$ from the spatial guidance image $\mathbf{G}_{\text{HR}}$. This guided deep decoder proposed in [28] is designed to span the space of the representation coefficients $\mathbf{A}$ from a low-dimensional manifold latent variables $\mathbf{Z}$ with the spatial prior information encoded by their parameters.

The network takes as inputs a randomly generated noise $\mathbf{Z}$ of size equal to the size of the representation matrix $\mathbf{A}$ and the high spatial resolution image $\mathbf{Y}_{\text{HR}}$ as the guidance image $\mathbf{G}_{\text{HR}}$. The network consists of two streams. The first one is a U-net based encoder-decoder architecture while the second one is a deep decoder and comprises upsampling refinement

Algorithm 1: Multiband image inversion: generic formulation

Input: Set $\mathbf{Y}$ of observed images, regularization parameters λ and μ.

1: Identify a basis $\mathbf{V}$ of the spectral subspace using PCA.
2: Train the deep generative decoder $\mathsf{D}_\theta(\cdot)$.
3: Initialization: $\mathbf{A}$, $\mathbf{Z}$ and $\mathbf{U}$ with zeros.
4: **while** not converged **do**:
5: Update $\mathbf{A}$ using (13);
6: Update $\mathbf{Z}$ using (14);
7: Update $\mathbf{U}$ using (15);
8: **end while**
9: $\hat{\mathbf{X}} = \mathbf{V}\mathsf{D}_\theta(\mathbf{Z})$.

Output: Estimated multiband image $\hat{\mathbf{X}}$.

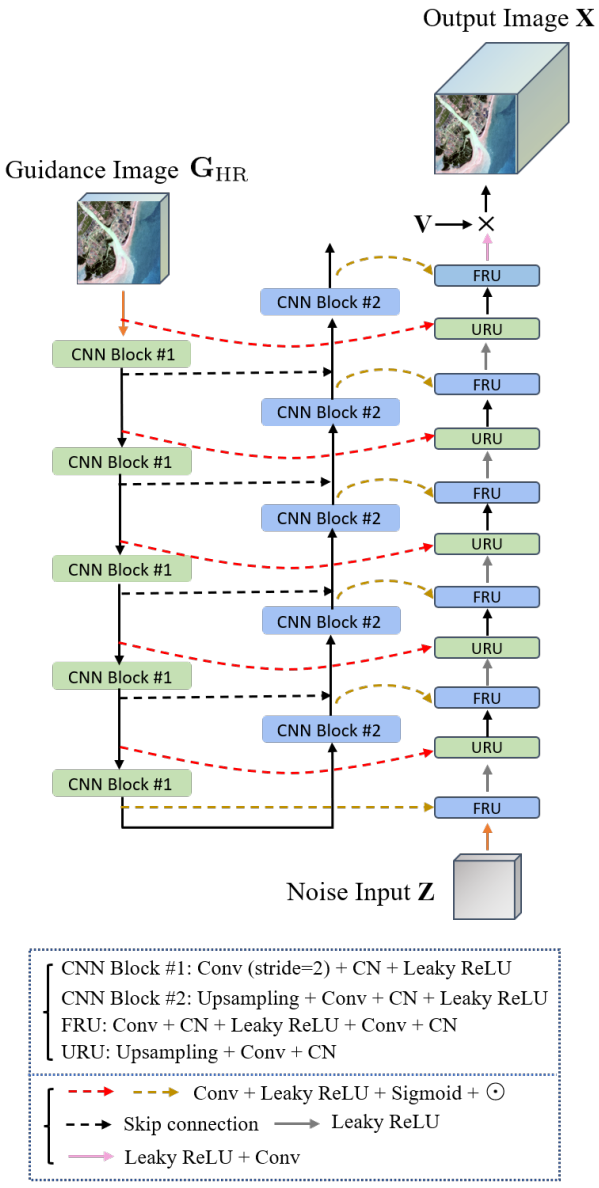

Fig. 1. The architecture of the guided deep generative model.

units (URU) and feature refinement units (FRU). The whole architecture is depicted in Fig. 1. Inspired by the design of deep image prior architectures [27], the deep decoder is trained to map a random generated noise $\mathbf{Z}$ to the estimated subspace coefficients $\mathbf{A} = \mathsf{D}_\theta(\mathbf{Z})$. The image structure information is encoded in the network parameters which can be resorted as an implicit image prior. The particularity of the GDD is the following: the input of the encoder-decoder network is assumed to be the auxiliary image of highest spatial resolution. It plays the role of guidance image whose spatial features are extracted at different scales to be used as conditional weights to guide the deep decoder. This model is trained as recommended by the authors in [28] by minimizing the mean square error between the target measurement $\mathcal{H}(\mathbf{Y})$ and the reconstructed image $\mathbf{V}\mathsf{D}_\theta(\mathbf{Z})$ evaluated through the forward model $\mathcal{M}(\cdot)$

$$\mathcal{L}_{\text{GDD}} = \|\mathcal{H}(\mathbf{Y}) - \mathcal{M}(\mathbf{V}\mathsf{D}_\theta(\mathbf{Z}))\|^2_{\mathbf{\Sigma}^{-1}}. \tag{16}$$

Once trained, the decoder $\mathsf{D}_\theta(\cdot)$ can be used as the generator.

V. Applications

To demonstrate the versatility of the proposed method to tackle challenging multiband imaging inverse problems, it is instantiated for two ubiquitous tasks, namely image fusion and multiband image inpainting. The instances associated to these two applications are detailed in what follows.

A. Multiband image fusion

1) Problem formulation: Given a pair of observed images $\mathbf{Y} = \{\mathbf{Y}_{\text{HR}}, \mathbf{Y}_{\text{HS}}\}$ of complementary resolutions, multiband image fusion aims to fuse a high spatial and low spectral resolution image $\mathbf{Y}_{\text{HR}}$ with a low spatial resolution and high spectral resolution image $\mathbf{Y}_{\text{HS}}$. The fused image $\mathbf{X}$ is expected to be of the highest spatial and spectral resolutions, $B = B_{\text{HS}}$ and $N = N_{\text{HR}}$. When the high spatial resolution image is a panchromatic image $\mathbf{Y}_{\text{HR}}$ (i.e., $B_{\text{HR}} = 1$), typical scenarios of this problem are referred to as pansharpening or hyperspectral pansharpening if the complementary image is a multispectral or hyperspectral images, respectively. The generic model (1) can then be instantiated by specifying the operators $\mathcal{H}(\cdot)$ and $\mathcal{M}(\cdot)$ as

$$\mathcal{H}(\mathbf{Y}) = \begin{bmatrix} \mathbf{Y}_{\text{HS}} \\ \mathbf{Y}_{\text{HR}} \end{bmatrix} \text{ and } \mathcal{M}(\mathbf{X}) = \begin{bmatrix} \mathbf{XBS} \\ \mathbf{RX} \end{bmatrix} \tag{17}$$

where $\mathbf{B} \in \mathbb{R}^{N \times N}$ is a cyclic convolution operator which stands for a spatial blurring, $\mathbf{S}$ denotes a regular spatial downsampling matrix with a downsampling factor denoted by f and $\mathbf{R} \in \mathbb{R}^{\tilde{B} \times B}$ is a spectral response. In most works of the literature [1], [3], the spectral regularization (7) is designed after conducting a principal component analysis of the high spectral resolution image, i.e. the spectral guidance image is chosen as $\mathbf{G}_{\text{HS}} = \mathbf{Y}_{\text{HS}}$. Conversely, the spatial regularization (3) is generally chosen as a parametric model promoting expected spatial characteristics, e.g., total variation [5] or Sobolev [29] for promoting piece-wise constant or smooth patterns, respectively. Instead, we propose to explicitly spatially regularize the fusion problem by resorting to the

high spatial resolution image as a spatial guidance image, i.e., $\mathbf{G}_{\text{HR}} = \mathbf{Y}_{\text{HR}}$. The covariance matrix $\mathbf{\Sigma}$ is assumed to specify sensor independent, spatially and spectrally white noises, which allows the optimization problem (11) to be finally rewritten as

$$\min_{\mathbf{Z}} \frac{1}{\sigma_{\text{HS}}^2}\|\mathbf{Y}_{\text{HS}} - \mathbf{V}\mathsf{D}_\theta(\mathbf{Z})\mathbf{B}\mathbf{S}\|_{\text{F}}^2 + \frac{1}{\sigma_{\text{HR}}^2}\|\mathbf{Y}_{\text{HR}} - \mathbf{R}\mathbf{V}\mathsf{D}_\theta(\mathbf{Z})\|_{\text{F}}^2 + \lambda\|\mathbf{Z}\|_{\text{F}}^2.$$

The generic algorithm proposed in Section IV-B to solve (18) is instantiated below.

2) Optimization: Introducing $\mathbf{A} = \mathsf{D}_\theta(\mathbf{Z})$, the augmented Lagrangian function can be written as

$$\mathcal{L}(\mathbf{A},\mathbf{Z},\mathbf{U}) = \frac{1}{\sigma_{\text{HS}}^2}\|\mathbf{Y}_{\text{HS}} - \mathbf{V}\mathbf{A}\mathbf{B}\mathbf{S}\|_{\text{F}}^2 + \frac{1}{\sigma_{\text{HR}}^2}\|\mathbf{Y}_{\text{HR}} - \mathbf{R}\mathbf{V}\mathbf{A}\|_{\text{F}}^2 + \mu\left\|\mathsf{D}_\theta(\mathbf{Z}) - \mathbf{A} + \frac{\mathbf{U}}{2\mu}\right\|_{\text{F}}^2 + \lambda\|\mathbf{Z}\|_{\text{F}}^2. \tag{18}$$

Updating $\mathbf{A}$ according to the rule (13) is a strongly convex problem that can be solved analytically by forcing the corresponding gradient to be zero. An efficient implementation of the solution can be derived following the strategy proposed by [1]. By noting that $\mathbf{V}$ is an orthogonal matrix with $\mathbf{V}^\top\mathbf{V} = \mathbf{I}_{\widetilde{B}}$, it consists in solving the Sylvester equation

$$\mathbf{C}_1\mathbf{A} + \mathbf{A}\mathbf{C}_2 = \mathbf{C}_3 \tag{19}$$

with

$$\begin{aligned}\mathbf{C}_1 &= \frac{1}{\sigma_{\text{HR}}^2}(\mathbf{R}\mathbf{V})^\top(\mathbf{R}\mathbf{V}) + \mu\mathbf{I}_{\widetilde{B}}\\ \mathbf{C}_2 &= \frac{1}{\sigma_{\text{HS}}^2}(\mathbf{B}\mathbf{S})(\mathbf{B}\mathbf{S})^\top\\ \mathbf{C}_3 &= \frac{1}{\sigma_{\text{HS}}^2}\mathbf{V}^\top\mathbf{Y}_{\text{HS}}(\mathbf{B}\mathbf{S})^\top + \frac{1}{\sigma_{\text{HR}}^2}(\mathbf{R}\mathbf{V})^\top\mathbf{Y}_{\text{HR}} + \mu\left(\mathsf{D}_\theta(\mathbf{Z}) + \frac{\mathbf{U}}{2\mu}\right).\end{aligned} \tag{20}$$

The resulting algorithmic scheme is recalled in Algorithm 2 for completeness. Regarding the updating rules for $\mathbf{Z}$ and $\mathbf{U}$, they can follow the derivations in (14) and (15).

Algorithm 2: Solution for (19) by solving the Sylvester equation.

Input: $\mathbf{Y}_{\text{HR}}$, $\mathbf{Y}_{\text{HS}}$, $\mathbf{V}$, $\mathbf{B}$, $\mathbf{S}$, f, $\mathbf{R}$, $\mathsf{D}_\theta(\mathbf{Z})$, $\mathbf{U}$, σ_{HS}, σ_{HR} and μ.

1: Compute $\mathbf{C}_1$, $\mathbf{C}_2$ and $\mathbf{C}_3$ using (20).
2: Eigen-decomposition of $\mathbf{B}$: $\mathbf{B} = \mathbf{F}\mathbf{D}\mathbf{F}^{\text{H}}$.
3: $\overline{\mathbf{D}} = \mathbf{D}(\mathbf{1}_f \otimes \mathbf{I}_n)$.
4: Eigen-decomposition of $\mathbf{C}_1$: $\mathbf{C}_1 = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^{-1}$.
5: $\overline{\mathbf{C}}_3 = \mathbf{Q}^{-1}\mathbf{C}_3\mathbf{F}$.
6: **for** $i = 1$ to $\widetilde{B}$ **do**:
7: $\quad\overline{\mathbf{A}}_i = \lambda_i^{-1}\left(\overline{\mathbf{C}}_3\right)_i - \lambda_i^{-1}\left(\overline{\mathbf{C}}_3\right)_i\overline{\mathbf{D}}\left(\lambda_i f\mathbf{I}_n + \sum_{t=1}^f \mathbf{D}_t^2\right)\overline{\mathbf{D}}^{\text{H}}$
8: **end for**
9: $\widehat{\mathbf{A}} = \mathbf{Q}\mathbf{A}\mathbf{F}^{\text{H}}$.

Output: The estimate $\widehat{\mathbf{A}}$.

B. Multiband image inpainting

1) Problem formulation: Because of sensor malfunctions or miscalibrations, multiband images are often affected by so-called *dead pixels*, i.e., pixels with unreliable measurements. The full multiband image $\mathbf{X}$ should be restored from the degraded observation $\mathbf{\Omega}_{\text{b}}\mathbf{Y}_{\text{HS}}\mathbf{\Omega}_{\text{p}}$ where $\mathbf{\Omega}_{\text{b}} \in \{0,1\}^{\tilde{B}\times B}$ and $\mathbf{\Omega}_{\text{p}} \in \{0,1\}^{N\times\tilde{N}}$ stand for binary matrices acting as masks to identify the $\tilde{B}$ out of B non-corrupted bands and the $\tilde{N}$ out of N non-corrupted pixels $\tilde{N} \leq N$. This task is referred to as multiband image inpainting. In some applicative scenarios, an auxiliary image $\mathbf{Y}_{\text{HR}}$ of high spatial resolution can be easily acquired by a low cost tool of lower spectral resolution. For instance, in the context of hyperspectral imaging, this complementary image can be a RGB image ($\tilde{B} = 3$) [30]. For STEM, when dealing with electron energy loss spectroscopy (EELS), the acquisition of the EELS data can be easily preceded by annular dark-field (HAADF) imaging providing a single-band image ($\tilde{B} = 1$) with a full spatial resolution [2]. The set of available images is then $\mathbf{Y} = \{\mathbf{Y}_{\text{HS}}, \mathbf{Y}_{\text{HR}}\}$. The multiband image $\mathbf{Y}_{\text{HS}}$ and the auxiliary image $\mathbf{Y}_{\text{HR}}$ serve here as the spectral and spatial guidance images, respectively, i.e., $\mathbf{G}_{\text{HS}} = \mathbf{Y}_{\text{HS}}$ and $\mathbf{G}_{\text{HR}} = \mathbf{Y}_{\text{HR}}$. Then the generic model (1) can be instantiated by defining the operators $\mathcal{H}(\cdot)$ and $\mathcal{M}(\cdot)$ as

$$\mathcal{H}(\mathbf{Y}) = \mathbf{\Omega}_{\text{b}}\mathbf{Y}_{\text{HS}}\mathbf{\Omega}_{\text{p}} \text{ and } \mathcal{M}(\mathbf{X}) = \mathbf{\Omega}_{\text{b}}\mathbf{X}\mathbf{\Omega}_{\text{p}}. \tag{21}$$

The generic problem (11) can be rewritten as

$$\min_{\mathbf{Z}} \|\mathbf{\Omega}_{\text{b}}\mathbf{Y}_{\text{HS}}\mathbf{\Omega}_{\text{p}} - \Omega_{\text{b}}\mathbf{V}\mathsf{D}_\theta(\mathbf{Z})\mathbf{\Omega}_{\text{p}}\|_{\text{F}}^2 + \lambda\|\mathbf{Z}\|_{\text{F}}^2. \tag{22}$$

The next section details the corresponding instance of the generic algorithm proposed in Section IV-B to solve (22).

2) Optimization: The augmented Lagrangian associated with (22) writes

$$\mathcal{L}(\mathbf{A},\mathbf{Z},\mathbf{U}) = \|\mathbf{\Omega}_{\text{b}}\mathbf{Y}_{\text{HS}}\mathbf{\Omega}_{\text{b}} - \mathbf{\Omega}_{\text{b}}\mathbf{V}\mathbf{A}\mathbf{\Omega}_{\text{b}}\|_{\text{F}}^2 + \mu\left\|\mathsf{D}_\theta(\mathbf{Z}) - \mathbf{A} + \frac{\mathbf{U}}{2\mu}\right\|_{\text{F}}^2 + \lambda\|\mathbf{Z}\|_{\text{F}}^2, \tag{23}$$

where $\mathbf{A} = \mathsf{D}_\theta(\mathbf{Z})$ denotes the auxiliary variable, $\mathbf{U}$ is the Lagrangian multiplier. Updating the auxiliary variable $\mathbf{A}$ boils down to solving the quadratic problem

$$\min_{\mathbf{A}} \|\mathbf{\Omega}_{\text{b}}\mathbf{Y}_{\text{HS}}\mathbf{\Omega}_{\text{p}} - \mathbf{\Omega}_{\text{b}}\mathbf{V}\mathbf{A}\mathbf{\Omega}_{\text{b}}\|_{\text{F}}^2 + \mu\left\|\mathsf{D}_\theta(\mathbf{Z}) - \mathbf{A} + \frac{\mathbf{U}}{2\mu}\right\|_{\text{F}}^2. \tag{24}$$

Because of the large size of this problem, its resolution is not straightforward. Inspired by the fast implementations proposed in [31] and [20], the problem (24) can be rewritten by vectorizing all quantities. More precisely, by denoting $\mathbf{y}_{\text{HS}} = \text{vec}\{\mathbf{Y}_{\text{HS}}\}$, $\mathbf{a} = \text{vec}\{\mathbf{A}\}$, $\mathsf{d}(\mathbf{Z}) = \text{vec}\{\mathsf{D}_\theta(\mathbf{Z})\}$ and $\mathbf{u} = \text{vec}\{\mathbf{U}\}$ where $\text{vec}\{\cdot\}$ stacks the columns of the corresponding matrix, the problem is equivalent to

$$\min_{\mathbf{a}} \|\mathbf{M}\mathbf{y}_{\text{HS}} - \mathbf{M}\left(\mathbf{I}_B \otimes \mathbf{V}\right)\mathbf{a}\|_{\text{F}}^2 + \mu\left\|\mathsf{d}(\mathbf{Z}) - \mathbf{a} + \frac{\mathbf{u}}{2\mu}\right\|_{\text{F}}^2 \tag{25}$$

where $\mathbf{M} \in \mathbb{R}^{\tilde{B}N \times BN}$ is the vectorization-based counterpart binary matrix of the masks $\mathbf{\Omega}_{\rm b}$ and $\mathbf{\Omega}_{\rm p}$. It yields a closed-form solution of (24) given by

$$\mathbf{A} = \text{vec}^{-1}\left\{\left[\mathbf{Q}\mathbf{Q}^\top + \mu\mathbf{I}_{BN}\right]^{-1}\left[\mathbf{Q}\mathbf{M}\mathbf{y}_{\rm HS} + \mu\mathbf{g}\right]\right\} \quad (26)$$

where $\mathbf{Q} = \left(\mathbf{I}_B \otimes \mathbf{V}^\top\right)\mathbf{M}^\top$ and $\mathbf{g} = \mathsf{d}(\mathbf{Z}) + \mathbf{u}/2\mu$. The updates of $\mathbf{Z}$ and $\mathbf{U}$ are the same as in (14) and (15).

VI. Experiments

This section shows how the proposed framework performs when tackling multiband imaging problems detailed in Section V, namely fusion and inpainting. For each task, the performance of the proposed method, referred to as ADMM-GDD, is compared to the performances reached by dedicated state-of-the-art methods. These compared methods will be detailed in the respective sections (see Sections VI-B and VI-C, respectively). In addition, the proposed framework is instantiated when the generative model $\mathsf{D}_\theta(\cdot)$ is not spatially informed by the guidance image $\mathbf{G}_{\rm HR}$. To do so, the GDD detailed in Section IV-C is replaced by a variational autoencoder (VAE) trained on a generic data set. Considering this framework, referred to as ADMM-VAE, will allow to highlight the benefits of informing the generative model with the guidance image of high spatial resolution. Details regarding the architecture and the training of the VAE-based generative model are given in the Appendix. Finally, to evaluate the relevance of the splitting-based algorithm detailed in Section IV-B, Adam is used to directly solve (11) instead of implementing the ADMM. The corresponding methods are coined as Adam-GDD and Adam-VAE.

A. Quality Metrics

Five figures-of-merit are used to quantitatively compare the results provided by the algorithms.

- *PSNR*: The peak signal-to-noise ratio (PSNR) is used to quantitatively evaluate the global similarity between the ground-truth and the estimate. It consists in computing the average of single-band SNR over the bands. The bigger the better estimation.
- *SAM*: The spectral angle mapper (SAM) [32] is a spectral distortion metric. It is computed by averaging the SAM over the pixels. The smaller the better estimation.
- *UIQI*: The universal image quality index (UIQI) [33] evaluates the similarity of correlation, luminance and contrast. The single-band UIQI are averaged over the bands. The bigger the better estimation.
- *ERGAS*: The relative dimensionless global error in synthesis (ERGAS) [34] is a band-wise mean-normalized root-mean-square error (RMSE) which is expected to be robust to calibration. The overall ERGAS is averaged over the bands. The smaller the better estimation.
- *SSIM*: The structural similarity index (SSIM) [35] is widely used to measure the structural similarities of the gray image. It is extended to multiband image by averaging the band-wise SSIM over all bands. The bigger the better estimation.

(a) (b)

Fig. 2. Fusion experiment – Reference images: (a) Pavia University and (b) Moffett field.

TABLE I
Fusion experiment with the Pavia University data set – Quantitative results.

Methods	PSNR ↑	SAM ↓	UIQI ↑	ERGAS ↓	SSIM ↑
SLRP-DSP	32.0080	5.7446	0.9302	3.2892	0.8940
ZSL	31.6974	4.9137	0.9186	3.4616	0.8885
CMS	30.1149	7.6407	0.8809	4.0991	0.8239
Deep-HS-prior	32.0470	5.3685	0.9228	3.1756	0.8750
CNN-Fus	32.0671	5.7449	0.9302	3.2892	0.8940
GDD	33.4084	5.1652	0.9481	2.6745	0.9219
Adam-VAE	29.9202	5.4480	0.8746	4.3159	0.8150
ADMM-VAE	31.6041	5.2516	0.8951	4.1670	0.8530
Adam-GDD	33.8994	4.8336	0.9485	2.6741	0.9229
ADMM-GDD	**34.4054**	**4.4018**	**0.9547**	**2.5605**	**0.9285**

B. Multiband image fusion

Data – In this study, two simulated hyperspectral data sets, namely the Pavia University and Moffett field data sets, are used to evaluate the effectiveness of the proposed method when tackling a multiband image fusion problem (see Section V-A).

The Pavia University image was acquired over the urban area of Pavia University, Italy. It consists of 610×340 pixels ($N = 207400$) with $B = 93$ spectral bands after removing the water vapor absorption and noisy bands. A color composition of the image is depicted in Fig. 2(a). It is considered as the ground-truth reference image $\mathbf{X}$ of high spatial and high spectral resolutions to be recovered. The observed images have been synthetically generated from this reference image following a protocol similar to the one described in [1]. More precisely, a hyperspectral image $\mathbf{Y}_{\rm HS}$ of low spatial resolution is generated by applying a 5×5 Gaussian filter with a standard deviation set to 2 and then subsampling with a factor equal to 5 in the horizontal and vertical directions. A panchromatic image $\mathbf{Y}_{\rm HR}$ of high spatial resolution is obtained from the reference

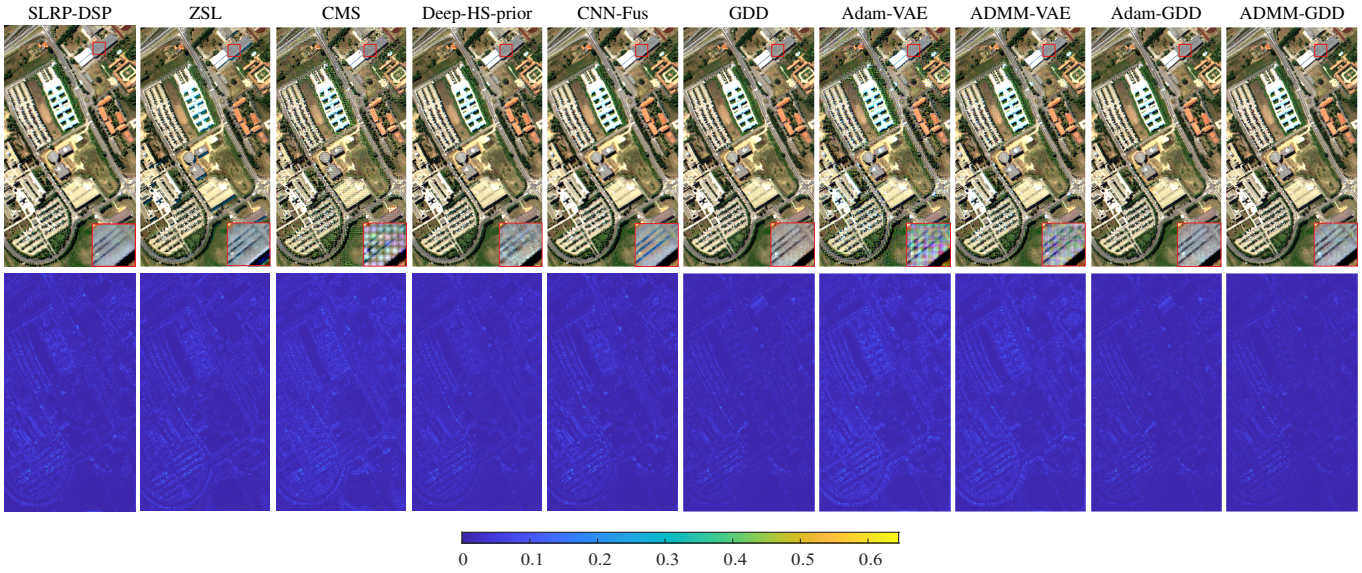

Fig. 3. Fusion experiment with the Pavia University data set – Color compositions of the fused image (1st row) and corresponding error images (2nd row).

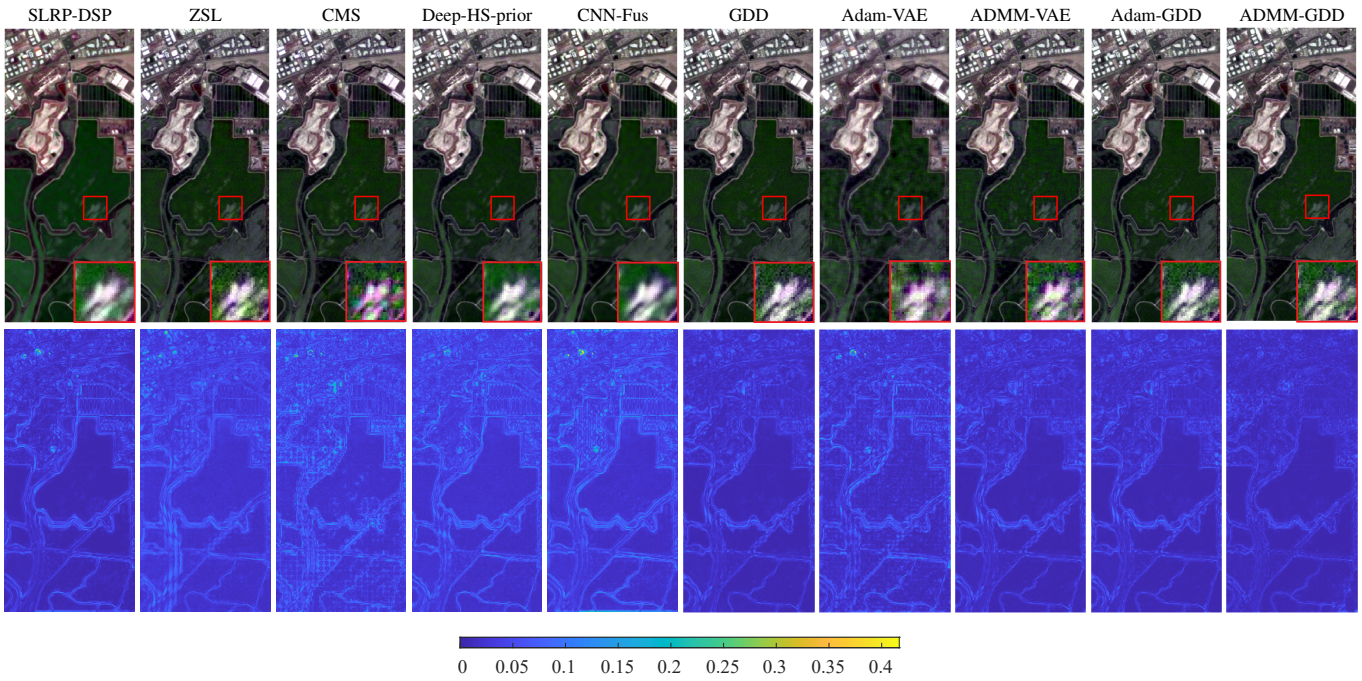

Fig. 4. Fusion experiment with the Moffett Field image data set – Color composition of the fused image (1st row) and corresponding error images (2nd row).

image by averaging all bands. Zero-mean additive Gaussian noises are added to the observed images with corresponding noise levels of SNR= 35dB for the hyperspectral image and SNR=30dB for the panchromatic image.

The Moffett field image was acquired by the JPL/NASA airborne visible/infrared imaging spectrometer (AVIRIS). This reference image is of size 396×184 pixels with $B = 176$ spectral bands after removing the water vapor absorption and noisy bands. Fig. 2(b) depicts a color composition of the image. As for the previous data set. To produce a hyperspectral image $\mathbf{Y}_{\mathrm{HS}}$ of lower spatial resolution, the reference image $\mathbf{X}$ has been spatially degraded by applying a 7×7 Gaussian filter with standard deviation set to 2 and then subsampling with a factor set to 7 in both directions. Conversely, the image $\mathbf{Y}_{\mathrm{HR}}$ has been obtained by averaging the first 41 bands of the full resolution image. This generation mimics the acquisition of the image $\mathbf{Y}_{\mathrm{HR}}$ by a panchromatic sensor with a limited spectral coverage. The two images have been corrupted by zero-mean additive Gaussian noises with SNR= 30dB for the hyperspectral image $\mathbf{Y}_{\mathrm{HS}}$ and SNR= 35dB for the panchromatic image $\mathbf{Y}_{\mathrm{HR}}$.

Compared methods – The proposed method has been compared with some recently proposed methods, including SLRP-DSP [36], ZSL [37], CMS [38], Deep-HS-prior [17] and CNN-Fus [4]. The GDD used to define the generative

TABLE II
FUSION EXPERIMENT WITH THE MOFFETT FIELD DATA SET – QUANTITATIVE RESULTS.

Methods	PSNR ↑	SAM ↓	UIQI ↑	ERGAS ↓	SSIM ↑
SLRP-DSP	33.6290	5.1002	0.7838	4.6536	0.7477
ZSL	34.4688	5.0912	0.8238	4.2792	0.8079
CMS	31.4534	5.2291	0.7570	4.8516	0.7183
Deep-HS-prior	33.4047	4.9510	0.7908	4.3346	0.7598
CNN-Fus	33.2445	5.0889	0.7653	4.8852	0.7257
GDD	34.6381	4.8182	0.8470	3.9694	0.8267
Adam-VAE	31.5450	5.7444	0.7627	5.8207	0.6782
ADMM-VAE	33.6141	5.0346	0.8102	4.9235	0.7986
Adam-GDD	34.8199	4.7849	0.8500	3.9512	0.8316
ADMM-GDD	**35.1060**	**4.7766**	**0.8636**	**3.6322**	**0.8410**

TABLE III
INPAINTING EXPERIMENT WITH THE UGR DATA SET – QUANTITATIVE RESULTS.

Methods	PSNR ↑	SAM ↓	UIQI ↑	ERGAS ↓	SSIM ↑
FastHyIn	38.3746	1.1329	0.8953	3.8298	0.9523
Deep-HS-prior	35.5531	1.4972	0.8413	5.0447	0.9165
wLRTR	38.5308	0.8827	0.916	3.1326	0.9679
PnP-In	38.6648	1.0307	0.9179	4.9074	0.9689
ADMM-ADAM	38.7166	0.9158	0.9191	3.3461	0.9700
GDD	37.1655	1.3830	0.8711	4.4531	0.9350
Adam-VAE	37.6023	1.1758	0.9083	4.0566	0.9439
ADMM-VAE	38.6609	1.1481	0.9181	3.2132	0.9695
Adam-GDD	37.6068	1.2814	0.8815	4.3924	0.9452
ADMM-GDD	**39.0391**	**0.8665**	**0.9204**	**2.9984**	**0.9706**

TABLE IV
INPAINTING EXPERIMENT WITH THE FRU DATA SET – QUANTITATIVE RESULTS.

Methods	PSNR ↑	SAM ↓	UIQI ↑	ERGAS ↓	SSIM ↑
FastHyIn	24.0655	26.3186	0.004	118.9453	0.2219
Deep-HS-prior	40.7871	1.0734	0.6267	21.8554	0.9507
wLRTR	25.2334	2.6928	0.0301	105.4779	0.4335
PnP-In	25.6018	16.8758	0.0223	105.9829	0.4128
ADMM-ADAM	40.7938	1.0684	0.6288	21.8406	0.9510
GDD	46.1837	0.9731	0.8011	16.9262	0.9868
Adam-VAE	27.4481	9.6011	0.1758	96.3161	0.5015
ADMM-VAE	32.2087	7.0052	0.2991	86.6622	0.6911
Adam-GDD	45.9297	0.9874	0.8032	17.0650	0.9871
ADMM-GDD	**46.8336**	**0.9657**	**0.8057**	**16.6714**	**0.9876**

model in the proposed framework is also considered as an end-to-end fusion method [28]. SLRP-DSP uses a self-supervised strategy to fine-tune the denoiser DnCNN to introduce spatial priors. ZSL uses zero-shot learning for fusion. This method has been implemented assuming the spectral and spatial responses are known. The CMS method excavates clustering manifold structure for super-resolution task, and the corresponding hyperparameters have been set to $\mu_{\text{CMS}} = 4 \times 10^{-4}$ and $\rho_{\text{CMS}} = 1.05$. Deep-HS-prior is a hyperspectral image enhancement method based on a deep image prior architecture. To train this model, the cost function is defined to ensure consistency of the pair of the hyperspectral image of low spatial resolution and the panchromatic image of high spatial resolution. The CNN-Fus is a plug-and-play based hyperspectral super-resolution method where a CNN-denoiser acts as a regularization. To train the VAE and GDD models, the number of epochs is set to 100 and 7000, and the learning rates of the Adam optimizer are fixed to 1×10^{-3} and 0.01, respectively. To train the VAE model, 5000 patches have been extracted from the guidance image and 10000 patches from an external image data set. The parameters of the proposed algorithm are set as $\mu = 7$, $\lambda = 0.7$ for Pavia University data set and $\mu = 2.3$, $\lambda = 0.23$ for Moffett Field data set.

Results – The quantitative results associated with the Pavia University and Moffett field data sets are reported in Table I and II, respectively, where the best results are highlighted in bold. Several important findings can be drawn for these results. Firstly, GDD competes favorably with respect to the compared model-based algorithm (CMS) and the four data-driven algorithms (SLRP-DSP, ZSL, Deep-HS-prior and CNN-Fus). These good results can be explained by the relevant deep architectures able to extract meaningful spatial information to guide the inversion process. In addition, when using the decoder as a guided generative model employed as a regularization, it provides even better performance. Indeed, the results obtained by Adam-GDD and ADMM-GDD demonstrate the relevance of the designed objective function (11). Secondly, the splitting-based minimization scheme ADMM-GDD proposed in Section IV-B provides significantly better results compared to a direct minimization by the Adam solver (Adam-GDD). This is also observed when the GDD-based regularization is replaced by a VAE model: ADMM-VAE performs better than Adam-VAE. Finally, the GDD regularization is shown to better capture the spatial information when compared to the VAE regularization, whatever the minimization technique (Adam or ADMM). This can be explained by the fact that GDD is guided by the sole guidance image of high spatial resolution while the VAE has been trained on an extended data set. In conclusion, the proposed ADMM-GDD method outperforms all compared methods since it combines the advantages of GDD as the regularization and ADMM as the minimization scheme. Fig. 3 and Fig. 4 illustrate the fusion results by depicting color compositions of the fused image recovered by the compared algorithms. It also depicts the spatial map of the pixel-wise reconstructed errors averaged over the spectral bands. It can be observed that the proposed method reconstructs more details and preserves their sharpness.

C. Multiband image inpainting

Data – This section reports experiments conducted to evaluate the performance of the proposed framework when tackling a multiband image inpainting problem (see Section V-B). Two acquisition scenarios are considered and are chosen to mimic

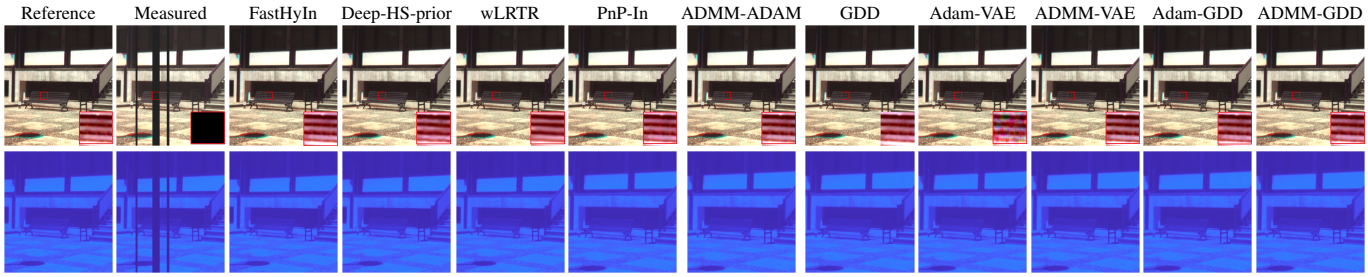

Fig. 5. Inpainting experiment with the UGR data set – Color compositions (1st row) and 47th band (2nd row) of the images. The dead (masked) pixels appear as vertical black lines in the measured images.

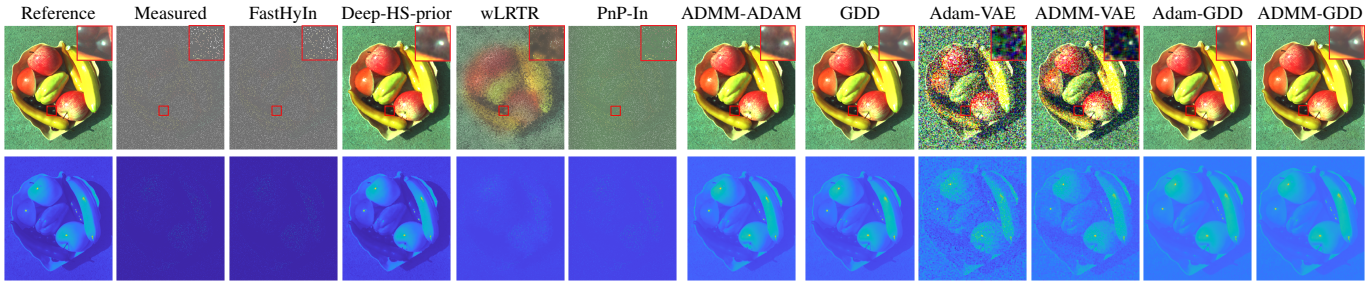

Fig. 6. Inpainting experiment with the Fru data set – Color compositions (1st row) and 120th band (2nd row) of the images. The non-acquired (masked) pixels appear in gray in the measured images.

archetypal applicative contexts. These scenarios rely on two distinct data sets.

The first data set, referred to as UGR data set in what follows, is selected from the UGR Hyperspectral Image Database. It contains pairs of hyperspectral and RGB images. The hyperspectral image $\mathbf{Y}_{\mathrm{HS}}$ is of size 1000×900 pixels with $B = 61$ spectral bands. It was acquired using the hyperspectral camera V-EOS by Photon etc and captured outdoor environments in Granada, Spain. The associated RGB image $\mathbf{Y}_{\mathrm{HR}}$ is of size $1000 \times 900 \times 3$ and is rendered by CIE Standard Illuminant D65 with the CIE 1931 2° Standard Observer with gamma correction ($\gamma = 0.6$). To mimic image corruption by stripe-like noise [39], we simulate dead pixels on randomly selected columns along specific bands. The masks $\mathbf{\Omega}_{\mathrm{b}}$ and $\mathbf{\Omega}_{\mathrm{p}}$ are defined such that 25 randomly selected bands are contaminated. A color composition of the image and a given band are depicted in Fig. 5 (1st column).

The second data set, termed as Fru data set, was acquired in our laboratory by a GaiaField camera. The image $\mathbf{Y}_{\mathrm{HS}}$ is of size 696×801 pixels with 256 spectral bands covering a spectral range from 400nm to 1000nm, with spectral resolution up to 0.58nm. An auxiliary RGB camera in the hyperspectral device was utilized to acquire the corresponding high resolution image $\mathbf{Y}_{\mathrm{HR}}$. In order to achieve a spatial partial acquisition, the bandwise mask $\mathbf{\Omega}_{\mathrm{b}}$ is chosen as $\mathbf{\Omega}_{\mathrm{b}} = \mathbf{I}_B$ ($B = \tilde{B}$) and the pixel-wise mask $\mathbf{\Omega}_{\mathrm{p}}$ is chosen to randomly select 5% of the spatial locations. In other words, only 5% of the spectra in $\mathbf{Y}_{\mathrm{HR}}$ are available to reconstruct the full image. This simulation protocol is similar to those followed in STEM-EELS where random sampling schemes are implemented to reduce the total acquisition time and preserve the observed samples [2]. The reference image and the corresponding measurements are shown in the first and second subimages of Fig. 6.

Compared methods – To evaluate the efficiency of the proposed framework, it is compared to several state-of-the-art algorithms, namely FastHyIn [31], Deep-HS-prior [17], weighted low-rank tensor recovery (wLRTR) [7], PnP-In [21], ADMM-ADAM [20] and GDD [28] as compared methods. The FastHyIn is a fast and competitive inpainting algorithm based on low-rank and sparse representations. The Deep-HS-prior is based on a deep image prior framework. The wLRTR is a unified low-rank tensor recovery method. The PnP-In is a recently proposed plug-and-play based inpainting method, which can plug denoiser priors for which we set $\rho = 1 \times 10^{-4}$ and directly use the pretrained Gray/RGB denoisers FFDNet [40] as the denoiser. The ADMM-ADAM solves the inpainting problem and integrates deep prior information. We use the restored results of Deep-HS-prior as the deep prior regularization with μ_a set to 1×10^{-3} and λ_a set to 0.01. Adam-VAE and Adam-GDD are also used as compared methods. To pretrain deep models embedded into the proposed framework, the number of epochs and the learning rate are set to 100 and 1×10^{-3} for the VAE and to 5000 and 0.01 for the GDD. As in the previous experiment, the VAE has been trained using 3000 patches from the guidance image and 10000 patches from an additional image data set, with patch size chosen as 25×25. The parameters μ and λ are set to 1×10^{-3} and 1×10^{-5}.

Results – The quantitative figures-of-merit PSNR, SAM, UIQI, ERGAS and SSIM obtained by the compared algorithms on the UGR data set are reported in Table III. All compared methods provide satisfactory reconstructed results. The proposed framework instantiated with an ADMM optimization

scheme and a GDD regularization outperforms all compared methods. The enhancement stems from two aspects: the first one comes from the guided prior learnt from the complementary RGB image, the other is the integrating use of convex optimization and deep generative priors. Fig. 5 shows the inpainting results for this data set. It is clear that the proposed method is closer to the ground-truth image. Table IV reports the quantitative results and Fig. 6 depicts the reconstructed images obtained on the Fru data set. For this challenging task, for which some spectra are completely unavailable for a large part of the spatial positions, conventional hyperspectral inpainting algorithms, such as FastHyIn, wLRTR and PnP-In, fail to recover the missing pixels and get bad restored results. Conversely, deep learning-based inpainting methods or approaches integrating deep networks provide good results. In particular the proposed method ADMM-GDD achieves the best results. It is worth noting that the result of ADMM-VAE is still blurry and noisy, which may due to the limited generative ability of the VAE model which does not fully exploit the guidance image $\mathbf{Y}_{\mathrm{HR}}$.

D. Computational times and convergence analysis

Experiments are conducted to evaluate the running times of the compared methods on the two considered tasks. Note that all the experiments are conducted on a server equipped with Intel Xeon E5-2650 v3 and an NVIDIA Tesla k80. CMS, CNN-Fus, FastHyIn, wLRTR, PnP-In, ADMM-ADAM are coded using MATLAB, and the remaining methods are implemented with Pytorch. The computational times required by the fusion (resp. inpainting) methods applied on the Pavia university (resp. UGR) data sets are reported in Table V. Deep-HS-prior and GDD are deep image prior based methods, i.e., their training directly solves the targeted inverse problems from one degraded image, which costs more time. SLRP-DSP, CNN-Fus, PnP-In, ADMM-ADAM and the proposed method are hybrid methods which merge physical-based and data-driven-based models. For these methods the time required to train the networks are not considered, since they depend on the deep architecture selected by the user. We can observe that the proposed method is an effective method for the two tasks. In particular the derived ADMM frameworks appear to significantly reduce the computational times with respect to the Adam-based solver.

This finding can be confirmed by assessing the convergence of the two optimization methods. Fig. 8 depicts the PSNRs of the reconstructed images as functions of the iteration number for ADMM-GDD and Adam-GDD. It can be seen that ADMM reaches higher PSNRs right after the 1st iteration and converges quickly. This also demonstrates that ADMM-GDD obtains better performance than Adam-GDD with less iterations and a lighter computational burden.

TABLE V
COMPUTATIONAL TIMES (S) OF COMPARED METHODS FOR THE FUSION EXPERIMENT (PAVIA UNIVERSITY DATA SET) AND THE INPAINTING EXPERIMENT (UGR DATA SETS).

	Methods	Time		Methods	Time
Fusion	SLRP-DSP	270.15	Inpainting	FastHyIn	355.08
	ZSL	0.80		Deep-HS-prior	2h43min
	CMS	232.11		wLRTR	10530.00
	Deep-HS-prior	6h27min		PnP-In	121.73
	CNN-Fus	85.76		ADMM-ADAM	108.91
	GDD	8h30min		GDD	43h14min
	Adam-VAE	1028.17		Adam-VAE	1295.20
	ADMM-VAE	30.11		ADMM-VAE	98.93
	Adam-GDD	1201.48		Adam-GDD	1319.21
	ADMM-GDD	34.49		ADMM-GDD	101.30

VII. CONCLUSION

This paper proposed a generic framework for multiband imaging inverse problems. It relies on a guided deep regularization designed to embed a generative prior learnt from an auxiliary acquisition of high spatial resolution. The resulting nonlinear objective function was minimized using an alternating direction method of multipliers. Contrary to a brute force method, i.e., based on automatic differentiation (e.g., Adam), this splitting based strategy had the great advantage of decomposing the initial problem into three simpler subproblems. In particular, for most of the inverse problems of interest, it was shown that one could resort to closed-form expressions of one subproblem. A nonlinear optimizer could be used for minimization the subproblem involving the deep regularization, which amounted to perform a nonlinear projection onto the range of the generative model. As a particular instance, the generative model was chosen as a guided deep decoder. However the proposed framework appeared to be sufficiently flexible to let the choice of the deep regularization to the end-user. The proposed framework was instantiated for two particular yet ubiquitous multiband imaging tasks, namely fusion and inpainting. Experiments conducted for these two tasks showed that the proposed framework outperformed state-of-the-art algorithms. Future work may include unrolling (or unfolding) the derived iterative optimization procedure to jointly learn the hyperparameters.

APPENDIX
VAE-BASED GENERATIVE MODEL

VAE has demonstrated excellent performance to model probability distributions of complex data sets and to generate new samples similar to the observations [41]. In this work, a VAE is used as a generative model and an isotropic Gaussian prior is typically imposed on the latent vectors $\mathbf{Z}$. As illustrated in Fig. 7, the architecture of this network consists of an encoder and a decoder. The encoder maps input image patches into the latent feature space $\mathcal{Z}$, and the decoder is trained to reconstruct the input image patches. The core of the network exploits convolutional layers. Apart from the input and output layers, the encoder and decoder consist of P and Q blocks, respectively. Each block is composed of a 3×3 convolution layer and a LeakyReLU activation function. In our work, the 2-dimensional input image patches used to train the

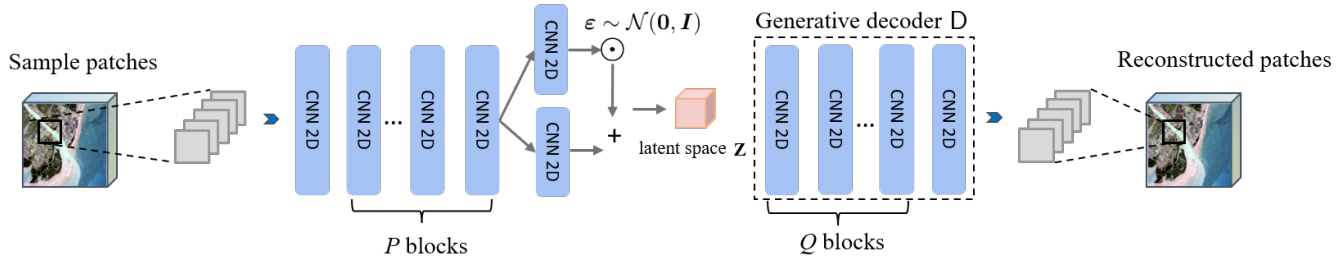

Fig. 7. The architecture of the VAE model.

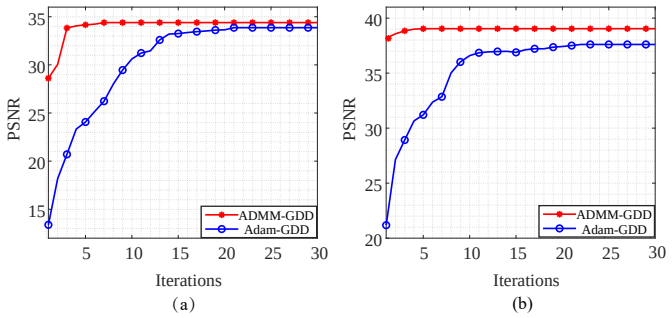

Fig. 8. PSNRs obtained by ADMM-GDD and Adam-GDD as functions of the iteration number: (a) fusion experiment on the Pavia University data set and (b) inpainting experiment on the UGR data set.

network are randomly selected from the guidance image and complementary images available from widely used image data sets such as DOTA [42] and HRSC2016 [43]. The objective function is then defined as

$$\begin{aligned}\mathcal{L}_{\text{VAE}} = &\frac{1}{N}\sum_{n=1}^{N}\|\widehat{\mathbf{x}}_n - \mathbf{x}_n\|_{\text{F}}^2 \\ &+ \lambda_{\text{KL}}\frac{1}{N}\sum_{n=1}^{N} D_{\text{KL}}\left(q\left(\boldsymbol{z}_n \mid \mathbf{x}_n\right)\|p(\boldsymbol{z})\right),\end{aligned} \tag{27}$$

where $\mathbf{x}_n$ and $\widehat{\mathbf{x}}_n$ denote the nth input and reconstructed patches, respectively. In (27), λ_{KL} is a hyperparameter adjusting the weight between the reconstruction error and the Kullback-Leibler (KL) divergence between the posterior and instrumental distributions. This network is trained thanks to the Adam optimizer. After training, the decoder is used as a generative model $\mathsf{D}_\theta(\cdot)$ which embeds the spatial regularization.

References

[1] Q. Wei, N. Dobigeon, and J.-Y. Tourneret, "Fast fusion of multi-band images based on solving a Sylvester equation," *IEEE Trans. Image Process.*, vol. 24, no. 11, pp. 4109–4121, 2015.

[2] E. Monier, T. Oberlin, N. Brun, M. Tencé, M. de Frutos, and N. Dobigeon, "Reconstruction of partially sampled multiband images—application to STEM-EELS imaging," *IEEE Trans. Comput. Imag.*, vol. 4, no. 4, pp. 585–598, 2018.

[3] Q. Wei, J. Bioucas-Dias, N. Dobigeon, and J.-Y. Tourneret, "Hyperspectral and multispectral image fusion based on a sparse representation," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 7, pp. 3658–3668, 2015.

[4] R. Dian, S. Li, and X. Kang, "Regularizing hyperspectral and multispectral image fusion by CNN denoiser," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 3, pp. 1124–1135, 2020.

[5] M. Simoes, J. Bioucas-Dias, L. B. Almeida, and J. Chanussot, "A convex formulation for hyperspectral image superresolution via subspace-based regularization," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 6, pp. 3373–3388, 2014.

[6] J. Xue, Y.-Q. Zhao, Y. Bu, W. Liao, J. C.-W. Chan, and W. Philips, "Spatial-spectral structured sparse low-rank representation for hyperspectral image super-resolution," *IEEE Trans. Image Process.*, vol. 30, pp. 3084–3097, 2021.

[7] Y. Chang, L. Yan, X.-L. Zhao, H. Fang, Z. Zhang, and S. Zhong, "Weighted low-rank tensor recovery for hyperspectral image restoration," *IEEE Trans Cybern.*, vol. 50, no. 11, pp. 4558–4572, 2020.

[8] Y. Chang, L. Yan, H. Fang, and C. Luo, "Anisotropic spectral-spatial total variation model for multispectral remote sensing image destriping," *IEEE Trans. Image Process.*, vol. 24, no. 6, pp. 1852–1866, 2015.

[9] Y. Wang, J. Peng, Q. Zhao, Y. Leung, X.-L. Zhao, and D. Meng, "Hyperspectral image restoration via total variation regularized low-rank tensor decomposition," *IEEE J. Sel. Top. Appl. Earth Observat. Remote Sens.*, vol. 11, no. 4, pp. 1227–1243, 2017.

[10] L. Wang, Z. Xiong, G. Shi, F. Wu, and W. Zeng, "Adaptive nonlocal sparse representation for dual-camera compressive hyperspectral imaging," *IEEE Tran. Patt. Analys. Machine Intell.*, vol. 39, no. 10, pp. 2104–2111, 2016.

[11] Y.-R. Fan and T.-Z. Huang, "Hyperspectral image restoration via superpixel segmentation of smooth band," *Neurocomputing*, vol. 455, pp. 340–352, 2021.

[12] W. Dong, H. Wang, F. Wu, G. Shi, and X. Li, "Deep spatial–spectral representation learning for hyperspectral image denoising," *IEEE Trans. Comput. Imag.*, vol. 5, no. 4, pp. 635–648, 2019.

[13] R. Wong, Z. Zhang, Y. Wang, F. Chen, and D. Zeng, "HSI-IPNet: Hyperspectral imagery inpainting by deep learning with adaptive spectral extraction," *IEEE J. Sel. Top. Appl. Earth Observat. Remote Sens.*, vol. 13, pp. 4369–4380, 2020.

[14] Q. Liu, H. Zhou, Q. Xu, X. Liu, and Y. Wang, "PSGAN: A generative adversarial network for remote sensing image pan-sharpening," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 12, pp. 10 227–10 242, 2020.

[15] Y. Qian, H. Zhu, L. Chen, and J. Zhou, "Hyperspectral image restoration with self-supervised learning: A two-stage training approach," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–17, 2021.

[16] L. Zhang, J. Nie, W. Wei, Y. Li, and Y. Zhang, "Deep blind hyperspectral image super-resolution," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 6, pp. 2388–2400, 2020.

[17] O. Sidorov and J. Yngve Hardeberg, "Deep hyperspectral prior: Single-image denoising, inpainting, super-resolution," in *Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops*, 2019, pp. 0–0.

[18] R. Dian, S. Li, A. Guo, and L. Fang, "Deep hyperspectral image sharpening," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 11, pp. 5345–5355, 2018.

[19] X. Wang, J. Chen, Q. Wei, and C. Richard, "Hyperspectral image super-resolution via deep prior regularization with parameter estimation," *IEEE Trans. Circuits Syst, Video Technol.*, vol. 32, no. 4, pp. 1708–1723, 2021.

[20] C.-H. Lin, Y.-C. Lin, and P.-W. Tang, "ADMM-ADAM: A new inverse imaging framework blending the advantages of convex optimization and deep learning," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–16, 2021.

[21] Z. Lai, K. Wei, and Y. Fu, "Deep plug-and-play prior for hyperspectral image restoration," *Neurocomputing*, vol. 481, pp. 281–293, 2022.

[22] J. M. Ramirez, J. I. Martínez-Torre, and H. Arguello, "LADMM-Net: An unrolled deep network for spectral image fusion from compressive data," *Signal Processing*, vol. 189, p. 108239, 2021.

[23] Q. Xie, M. Zhou, Q. Zhao, Z. Xu, and D. Meng, "MHF-Net: An interpretable deep network for multispectral and hyperspectral image fusion," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 3, pp. 1457–1473, 2022.

[24] J. M. Bioucas-Dias and J. M. Nascimento, "Hyperspectral subspace identification," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 8, pp. 2435–2445, 2008.

[25] A. A. Green, M. Berman, P. Switzer, and M. D. Craig, "A transformation for ordering multispectral data in terms of image quality with implications for noise removal," *IEEE Trans. Geosci. Remote Sens.*, vol. 26, no. 1, pp. 65–74, 1988.

[26] C. G. Turhan and H. S. Bilge, "Recent trends in deep generative models: a review," in *2018 3rd International Conference on Computer Science and Engineering (UBMK)*. IEEE, 2018, pp. 574–579.

[27] D. Ulyanov, A. Vedaldi, and V. Lempitsky, "Deep image prior," in *Proc. IEEE Conf. Computer vision and Pattern Recognition (CVPR)*, 2018, pp. 9446–9454.

[28] T. Uezato, D. Hong, N. Yokoya, and W. He, "Guided deep decoder: Unsupervised image pair fusion," in *European Conference on Computer Vision*. Springer, 2020, pp. 87–102.

[29] C. Guilloteau, T. Oberlin, O. Berné, and N. Dobigeon, "Hyperspectral and multispectral image fusion under spectrally varying spatial blurs–application to high dimensional infrared astronomical imaging," *IEEE Trans. Comput. Imag.*, vol. 6, pp. 1362–1374, 2020.

[30] L. Yan, X. Wang, M. Zhao, M. Kaloorazi, J. Chen, and S. Rahardja, "Reconstruction of hyperspectral data from RGB images with prior category information," *IEEE Trans. Comput. Imag.*, vol. 6, pp. 1070–1081, 2020.

[31] L. Zhuang and J. M. Bioucas-Dias, "Fast hyperspectral image denoising and inpainting based on low-rank and sparse representations," *IEEE J. Sel. Top. Appl. Earth Observat. Remote Sens.*, vol. 11, no. 3, pp. 730–742, 2018.

[32] R. H. Yuhas, A. F. Goetz, and J. W. Boardman, "Discrimination among semi-arid landscape endmembers using the spectral angle mapper (SAM) algorithm," in *JPL, Summaries of the Third Annual JPL Airborne Geoscience Workshop. Volume 1: AVIRIS Workshop*, 1992.

[33] Z. Wang and A. C. Bovik, "A universal image quality index," *IEEE Signal Proc. Lett.*, vol. 9, no. 3, pp. 81–84, 2002.

[34] L. Wald, "Quality of high resolution synthesised images: Is there a simple criterion?" in *Third conference" Fusion of Earth data: merging point measurements, raster maps and remotely sensed images"*. SEE/URISCA, 2000, pp. 99–103.

[35] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, 2004.

[36] Q. Zhang, Q. Yuan, M. Song, H. Yu, and L. Zhang, "Cooperated spectral low-rankness prior and deep spatial prior for HSI unsupervised denoising," *IEEE Trans. Image Process.*, vol. 31, pp. 6356–6368, 2022.

[37] R. Dian, A. Guo, and S. Li, "Zero-Shot Hyperspectral Sharpening," *IEEE Trans. Pattern Anal. Mach. Intell.*, 2023.

[38] L. Zhang, W. Wei, C. Bai, Y. Gao, and Y. Zhang, "Exploiting clustering manifold structure for hyperspectral imagery super-resolution," *IEEE Trans. Image Process.*, vol. 27, no. 12, pp. 5969–5982, 2018.

[39] P. M. Mather and M. Koch, *Computer processing of remotely-sensed images: an introduction*. John Wiley & Sons, 2011.

[40] K. Zhang, W. Zuo, and L. Zhang, "FFDNet: Toward a fast and flexible solution for cnn-based image denoising," *IEEE Trans. Image Process.*, vol. 27, no. 9, pp. 4608–4622, 2018.

[41] D. P. Kingma and M. Welling, "Auto-encoding variational Bayes," *arXiv preprint arXiv:1312.6114*, 2013.

[42] G.-S. Xia, X. Bai, J. Ding, Z. Zhu, S. Belongie, J. Luo, M. Datcu, M. Pelillo, and L. Zhang, "DOTA: A large-scale dataset for object detection in aerial images," in *Proc. IEEE Conf. Computer vision and Pattern Recognition (CVPR)*, 2018, pp. 3974–3983.

[43] H. Su, S. Wei, S. Liu, J. Liang, C. Wang, J. Shi, and X. Zhang, "HQ-ISNet: High-quality instance segmentation for remote sensing imagery," *Remote Sens.*, vol. 12, no. 6, p. 989, 2020.