Bayesian Approach in a Learning-Based Hyperspectral Image Denoising Framework

Hazique Aetesam, Suman K Maji, Hussein Yahia

HAL Id: hal-03502378
https://inria.hal.science/hal-03502378

Submitted on 24 Dec 2021

Bayesian Approach in a Learning-Based Hyperspectral Image Denoising Framework

HAZIQUE AETESAM[1], **SUMAN KUMAR MAJI**[1], **AND HUSSEIN YAHIA**[2], (Member, IEEE)

[1]Department of Computer Science and Engineering, Indian Institute of Technology Patna, Patna 801106, India
[2]Geostat Team, INRIA Bordeaux, 33405 Talence, France

Corresponding authors: Hazique Aetesam (hazique.pcs16@iitp.ac.in) and Suman Kumar Maji (smaji@iitp.ac.in)

ABSTRACT Hyperspectral images (HSI) are corrupted by a combination of Gaussian and impulse noise. Successful denoising of HSI data increases the accuracy of high-level vision operations like classification, target tracking and land-cover problem. On the one hand, the traditional approach of handling the denoising problem using maximum a posteriori (MAP) criterion is often restricted by the time-consuming iterative optimization process and design of hand-crafted priors to obtain an optimal result. On the other hand, the discriminative learning-based approaches offer fast inference speed over a trained model; but are highly sensitive to the noise level used for training. A discriminative model trained with a loss function which does not accord with the Bayesian degradation process often leads to sub-optimal results. In this paper, we design the training paradigm emphasizing the role of loss functions in neural network; similar to as observed in model-based optimization methods. Further, Bayesian motivated loss functions also act as priors to constrain the solution space to the types of noise observed in hyperspectral image acquisition process. As a result, loss functions derived in Bayesian setting and employed in neural network training boosts the denoising performance. Extensive analysis and experimental results on synthetically corrupted and real hyperspectral datasets suggest the potential applicability of the proposed technique under a wide range of homogeneous and heterogeneous noisy settings. Classification results over the real dataset and associated metrics like kappa coefficient and overall accuracy used as the task-based evaluators further support our hypothesis made in the proposed methodology.

INDEX TERMS Bayesian estimation, discriminative learning, Gaussian-impulse noise, hyperspectral imaging, residual learning.

I. INTRODUCTION

Images acquired from hyperspectral sensors comprise more than 100 narrow spectral bands over a wide electromagnetic spectrum range of 400 nm to 2500 nm. This facilitates detailed acquisition of the target area which is not possible using conventional RGB cameras. Hyperspectral imaging (HSI) finds its applications in areas involving remote sensing, agriculture [1], medicine [2], [3], military, pharmaceuticals [4]; to name a few. However, there are some inherent limitations in the acquisition equipment which result in noise, horizontal/vertical stripes and dead pixels. This happens as a result of fluctuations in power supply, dark current, calibration error and sensor sensitivity; among other factors. Thermal and shot noise corrupt the image due to

The associate editor coordinating the review of this manuscript and approving it for publication was Wentao Fan.

limited energy captured by narrow spectral bands [5]. Cumulative effects of these factors render noise in acquired HSI data which is a combination of Gaussian and impulse noise. Consequently, noise removal is a necessary preprocessing step that affects the subsequent post-processing tasks like classification, object tracking and target detection [6].

II. RELATED WORKS

Denoising methods in HSI can be classified into three different types: filtering-based, statistical (variational) techniques and learning-based methods. A summary of different categories of denoising methods is tabulated in Table 1. Filtering-based methods can work in either spatial or transform domain and statistical approaches can be combined with low-rank based methods. Therefore, some of the methods discussed in subsequent paragraphs share more than one category (see Table 1). This paragraph discusses filter-based methods.

TABLE 1. A summary of HSI denoising strategies.

Filtering-based (spatial domain)	[7], [8]
Filtering-based (transform domain)	[9], [10]
Statistical (Variational) approaches	[11], [12], [13], [14], [15], [16], [17]
TV-based methods	[18], [19], [15]
Low-rank methods	[20], [6], [18], [21], [19], [22], [23], [24], [24]
Learning-based approaches	[25], [26], [27], [28], [29], [30], [31], [32]

Statistical and learning-based techniques are deferred till the next two paragraphs. Authors in [7] proposed two different strategies under two different scenarios: modified version of median filtering when all channels are noisy and spectrum smoothing when only specific channels are noisy. However, choosing a fixed threshold is a not a feasible solution when noise components themselves can be variable. A non-local means (NLM) filtering approach adapted to 3D HSI data and signal-dependent nature of noise is proposed in [8]. In the same year, assuming that the underlying noise distribution is a combination of signal dependent Poisson and signal independent Gaussian components; Minchao *et al.* [9] used discrete cosine transform (DCT) as the sparsifying basis after performing variance stabilization. However, all these three methods [7]–[9] did not take into account heavy-tailed distribution of noise introduced during transmission of remotely sensed data in HSI signals; i.e., impulse noise. In another work, dissimilarity in signal regularity in terms of variable spectral noise levels is explored in [10] using shrinkage of wavelet coefficients under derivative domain. However, in [10], since all experiments are performed over real HSI data, there is no reliable quantitative evaluation using popular full-reference image quality metrics like peak signal-to-noise ratio (PSNR) between denoised and noiseless groundtruth image. In general, we can conclude that the performance of filtering-based methods are restricted by the size of smoothing parameter of the filter as in [7] and shrinkage thresholding level for the wavelet coefficients as in [10]. All these decisions require manual intervention and hence not practically realizable at large scale.

Statistical approaches work in two different yet intertwined ways: noise [11] and data modelling. A rank-1 tensor decomposition using eigenvalue intensity sorting technique is detailed in the work of Guo *et al.* [20]. A variational framework for sparse recovery is designed by the authors in [12] with consideration only to Gaussian noise in HSI data while only impulse noise is considered in [13]. However, both Gaussian and impulse noise is considered while modelling the optimization framework in [14] and [15]. All the three variational problems [12], [14], [15] are solved using Split-Bregman as the splitting technique to divide a complex optimization problem into smaller units. A 3D modelling of composite noisy HSI data followed by its optimization using primal-dual hybrid gradient (PDHG) is proposed in [16].

A two-phase splitting approach for the recovery of clean estimation is discussed in [17]. Low-rank assumption based on spatial and spectral similarity of data is explored in multitude of works. Also, it has been observed that total variation (TV) as image prior in varied forms is investigated along with low-rank assumption in HSI data. Authors in [6] lexicographically ordered 3D cube in 2D matrix based on the assumption that clean HSI data is low-rank. In a method proposed by He *et al.* [18], authors proposed a three-fold approach for the recovery of clean data. This includes the application of nuclear norm for exploiting spectral low-rank property and TV regularization for denoising piecewise-smooth regions in the spatial bands. Further ℓ_1-norm is utilized for the removal of sparse noise, stripes and dead pixels. However, both these methods [6], [18] consider similarity only in the spectral dimension but pixels in each layer can themselves be of low-rank. Local redundancy and correlation (RAC) in the spectral dimension and global RAC in the spatial dimension is employed in [21] for modeling the similarity among data pixels. Further, denoising is performed by sparse approximation of learned dictionary. A spatial-spectral TV regularised global smoothness constraint and local low-rank components are separated in [19]. The resultant optimization problem is solved using augmented Lagrange multiplier method. An approach based on non-local self-similarity and sparse representation is discussed in [22]. Most of the previously discussed methods did not take into account computational burden encountered in low-rank based methods using singular value decomposition (SVD). Authors in [23] mitigate this issue by requiring only upper bound of matrix rank in the associated principal component analysis (PCA). Authors in [24] claim that the present low-rank based methods do not consider spectral continuity among bands. As a result, a double-factor regularised low-rank matrix factorization (LFTR-DFR) approach is proposed in [24]. The resultant optimization setup is solved using proximal alternating minimization method. In general, we can conclude that 2D low-rank based methods cause large spectral distortion along the channel dimension. This issue becomes less worse with tensor-based optimization methods; but the resultant computational cost is severe.

Limited by the manual tuning of filtering parameters in filter-based approaches and hand-crafted priors and iterative design strategy in statistical approaches, there is a third category; which is essentially data-driven. It tries to overcome the stated limitations. More precisely, these data-driven approaches estimate true data from its corrupted version by learning a large set of parameters over input-target datasets. In recent works, it is observed that learning-based techniques are utilized for noise reduction in HSI data. Some notable contributions include [25]–[28], [29] and [30]. Authors in [25] employ cubic noisy-clean image pairs in their learning approach where non-linearity functions are also trainable along with convolutional weights and biases. A generative model to HSI denoising is employed in [28]. To exploit multi-level and multi-scale feature extraction from

HSI data, residual learning strategy is employed in [30] by concatenating features learned from several intermediate layers into one. In another approach, a single model is designed to accommodate noisy data from several noise levels [29]. Here, noise level estimation map is fed as prior to a deep convolutional neural network (CNN) for clean data estimation. Inspired by the pioneering work of Deep Image Prior (DIP); noisy HSI data is recovered in [27] using the assumption of prior introduced by untrained neural network and early stopping as the regularization strategy. A self-supervised learning approach using separable image prior is explored in [26]. However, since most of these models [25]–[29] are not trained over mixed noise in corrupted HSI data; they yield sub-optimal results when these techniques are tested over real HSI data. However, there are some notable works [31], [32] which do consider mixed-case scenario. For training the residual learning model, authors in [31] fed image gradients in the spatial and adjacent spectral dimensions to consider the effect of noise in neighbouring pixels. In [32], a convolutional neural network (CNN) is modelled as constraint non-negative matrix factorization strategy.

Taking the benefits of fast testing speed of trained neural networks and prior assumption of image degradation in a statistical Bayesian setting using a model-driven strategy; in this paper, we design a training approach which synchronises with the image degradation observed in HSI data and prior introduced by weights learning in the end-to-end training of noisy-clean image pairs. The paper initiates with the preliminary information and motivation in section III. Proposed work is discussed in section IV. Experimental results are presented in section V. Finally, the paper concludes in section VI.

III. PRELIMINARY AND MOTIVATION

Image formation model in HSI data is given by [6], [18]:

$$h = f + g + s \tag{1}$$

where clean data $f \in R^{m \times n \times l}$ is corrupted by additive Gaussian noise approximated by normal distribution $g \sim \mathcal{N}(0, \sigma_g^2)$ with zero mean and standard deviation σ_g. This Gaussian corrupted data $(f + g)$ is further degraded by sparse/random-valued impulse noise s (which heavily corrupts limited number of pixels); approximated by Laplacian distribution [33], [34] $s \sim \mathcal{L}(0, \sigma_s)$ with $\mu_s = 0$ as the location parameter and σ_s as the scaling parameter. Higher values of σ_g^2 and σ_s denote higher noise levels. Collective corruption by Gaussian component g and impulse component s generates noisy observation h. Dimensions of all the variables $\{h, g, s\}$ are same as f.

From the Bayesian point of view, estimation of clean data $\hat{f} \simeq f$ can be obtained by solving the following risk minimization task according to Maximum a posteriori (MAP) criterion using an optimization technique.

$$\hat{f} = \arg\min_{f} -\log\left[p(f|h)\right] = \arg\min_{f} -\log[p(h|f) \cdot p(f)] \tag{2}$$

where $p(f|h)$ is the posterior distribution, $p(h|f)$ is the likelihood term which takes appropriate form based on the distribution used to model the degradation process and $p(f)$ is the prior enforcing required properties that we intend to achieve in the solution space. Deciding an appropriate prior requires extensive knowledge of the working domain. In a learning-based paradigm, this extensive knowledge is obtained from large number of pairs of noisy-clean images.

In the recent past years, researchers have tried to establish links between model-driven optimization approaches and data-driven learning-based methods. Sparse coding using feed-forward learners [35], sparse outliers removal using ℓ_0-norm representations [36] and non-linear reactive diffusion models [37] are some representative works. More recently, authors in [38] have created an analogy between CNN and model-based approaches where direct inversion of the forward model followed by a CNN can be used in image inversion problems. In a typical multi-layer CNN, we consider that the output from the first layer is given as $f^1 = \mathcal{F}(W^{(1)} \otimes h + b^{(1)}) \in R^{m \times n \times B_1}$. Hence, output from the $d + 1^{\text{th}}$ layer is given by:

$$f^{d+1} = \mathcal{F}(W^{(d+1)} \otimes f^d + b^{(d+1)}) \tag{3}$$

where $f^{d+1} \in R^{m \times n \times B_{d+1}}$. Here, B_{d+1} is the number of filter outputs from the $(d + 1)^{\text{th}}$ layer. W, b and $\otimes$ are the weight matrices, biases and convolution operation respectively. $\mathcal{F}$ is a non-linear activation function like sigmoid or ReLU activation function. The expression above can be interpreted as the solution of an inverse problem through D number of layers when a corrupted input is fed into the model trained with corrupted-clean image pairs. In a typical optimization-based iterative approach, solution using gradient descent is given as:

$$f_{k+1} = f_k - \alpha \cdot \nabla F(f_k) \tag{4}$$

where f_k is the solution to the optimization problem obtained at k^{th} iteration and α is the step size. F is a multivariate differentiable function whose solution at successive steps descents towards a local minima. This iterative approach in Eq. (4) can be interpreted as D layers of a neural network (as in Eq. (3)) used at the inference time for predicting clean data from corrupted observations [28].

In a supervised discriminative learning environment, we basically minimize loss between the estimation $(\hat{f})$ and the ground truth (f) by optimizing a set of parameters θ through a learning process over N training pairs of noisy-clean data $\{h^{(i)}, f^{(i)}\}_{i=1}^N$ [39], [40].

$$\begin{cases} \min_{\theta} \mathcal{L}(\hat{f}, f) \\ \text{such that } \hat{f} = \arg\min_{f} \|h - f\|_p + \lambda \mathcal{R}(f; \theta) \end{cases} \tag{5}$$

where $\|\cdot\|_p$ is the l_p-norm whose exact form will be discussed in the proposed section (section IV). $\mathcal{R}(f)$ is the regularizer with the penalty term λ. The MAP guided learning process in Eq. (2) can be replaced by a non-linear function $\vartheta_\theta(h)$

parametrized by learning parameters θ [39] and a loss function $\mathcal{L}$ such that $\hat{f} = \vartheta_\theta(h)$. However, in our proposed work, we do not replace the MAP guided learning process with a non-linear function ϑ_θ but share this responsibility with the loss function used for model training.

We end this section with the following observation. At one extreme, we try to establish relationship between model and data-driven approaches. However, this relationship has to be reflected in our training process. Hence, in the following section, we train our model using carefully designed loss function that accords with the degradation process in the concerned imaging domain (here, HSI).

IV. THE PROPOSED DENOISING SOLUTION

With the aid of the likelihood term $-\log p(h|f)$ discussed in Eq. (2), if we consider that our observation is corrupted by noise following Gaussian distribution, the likelihood term takes the following form:

$$\underset{f}{\arg\min} - \log\left(\frac{1}{\sqrt{2\pi\sigma_g^2}}\exp\left(-\frac{(h-f)^2}{2\sigma_g^2}\right)\right) \quad (6a)$$

Eliminating the constant terms and solving for data to be estimated (i.e., $\hat{f}$); Eq. (6a) is equivalent to minimizing the data-fitting term in ℓ_2-norm (also called Mean Square Error (MSE)).

$$\underset{f}{\arg\min}\frac{1}{2}\|h-f\|_2^2 \quad (6b)$$

Similarly, considering that impulse noise is modelled as Laplacian distribution [33], [34], the likelihood term takes the following expression:

$$\underset{f}{\arg\min} - \log\left(\frac{1}{2\sigma_s}\exp\left(-\frac{|h-f|}{\sigma_s}\right)\right) \quad (7a)$$

Minimizing the above expression while ignoring constant terms generate ℓ_1-norm [41]–[43] or mean absolute deviation (MAE) as the data-fitting term.

$$\underset{f}{\arg\min}\|h-f\|_1 \quad (7b)$$

Hence, we can directly observe that putting $p = 2$ and $p = 1$ (as discussed in Eq. (5) yields ℓ_2 (Eq. (6b)) and ℓ_1-norm (Eq. (7b)) respectively in the likelihood solution. As discussed in section I, noise in HSI data is the combined effect of two different sources; namely Gaussian and impulse noise handled using ℓ_2 and ℓ_1 terms respectively as the data fidelity terms. If a neural network (NN) $\vartheta_\theta(h)$ is used for learning parameters θ, the cumulative loss ($\ell_2 + \ell_1$) can be used to guide the learning process according to MAP estimator. Further, the convolutional nature of the NN is capable of learning complex prior information $p(f)$ in the data [40]. Hence, the collective loss is given by:

$$\mathcal{L}_\theta^{\ell_2+\ell_1} = \frac{1}{N}\sum_{i=1}^{N}\left[\frac{\alpha}{2}\|\vartheta_\theta(h_i)-f_i\|_2^2 + (1-\alpha)\,\|\vartheta_\theta(h_i)-f_i\|_1\right] \quad (8)$$

where N is the number of noisy-clean image pairs in a single batch; for model training using any stochastic gradient-based optimization method. α is the penalty term adjusting the relative contribution of constituent fitting terms. An ablation study is performed in section V-E confirming the rationale behind this formulation. If we further assume that ϑ_θ does not produce a direct mapping from noisy to clean data but residual image $v = \vartheta_\theta(h)$ where v can be predicted Gaussian/impulse noise as guided by ℓ_2/ℓ_1 loss respectively, then the clean estimation is represented by $\hat{f} = h - \vartheta_{\hat{\theta}}(h)$. To accommodate residual connection in our loss function, training process in Eq. (8) can be modified according to the following expression:

$$\mathcal{L}_\theta^{\ell_2+\ell_1} = \frac{1}{N}\sum_{i=1}^{N}\left[\frac{\alpha}{2}\|\vartheta_\theta(h_i)-r_i\|_2^2 + (1-\alpha)\|\vartheta_\theta(h_i)-r_i\|_1\right] \quad (9)$$

where $r = h - f$ is the true residual image and $\vartheta_\theta(h)$ is the residual image predicted by the network. Further, it is observed that the image representations captured by neural networks often ignore fine details as they are inherently suited for high-level vision tasks [44]. To constrain the solution space to retain those low-level statistics in the result, regularisers need to be incorporated in the objective function. Total variation (TV) is an excellent choice to retain sharp edges by capturing piece-wise constant features often washed out as a result of denoising process [44]. This regulariser can be expressed in its discrete form as finite difference approximation:

$$\mathcal{R}(f_i) = \sum_{p,q,r}((f_{p+1,q,r} - f_{p,q,r})^2 + (f_{p,q+1,r} - f_{p,q,r})^2$$
$$+ (f_{p,q,r+1} - f_{p,q,r})^2)^{\frac{1}{2}} \quad (10)$$

where $\{p, q, r\}$ are 3d image indices. Accommodating this regularizer in Eq. (9), we obtain the final loss as:

$$\mathcal{L}_\theta = \mathcal{L}_\theta^{\ell_2+\ell_1} + \lambda\mathcal{R}(f) \quad (11)$$

A. NETWORK ARCHITECTURE AND TRAINING OPTIONS

Architecture of the neural network (NN) used for training the model is shown in Fig. 1. Normal and dilated [45] 3D convolutions are used under two parallel paths. The size of filters is set at $5 \times 5 \times 5$. On the one hand, normal convolutions perform simple convolutions and cover a limited region of the receptive field. On the other hand, dilated convolutions with dilation factor 2 increase the receptive field to cover a wider area of the data under investigation. This helps in the extraction of multi-scale contextual knowledge from a wider patch of 3D data without adding any computational burden. Besides the extraction of multi-level features using two separate paths, the rationale behind the usage of a wider network rather than serially cascaded filters is motivated by the problem of vanishing gradient encountered in deeper networks [46]. As pointed out in [47], same set of features can be detected by increasing the number and size of convolutional
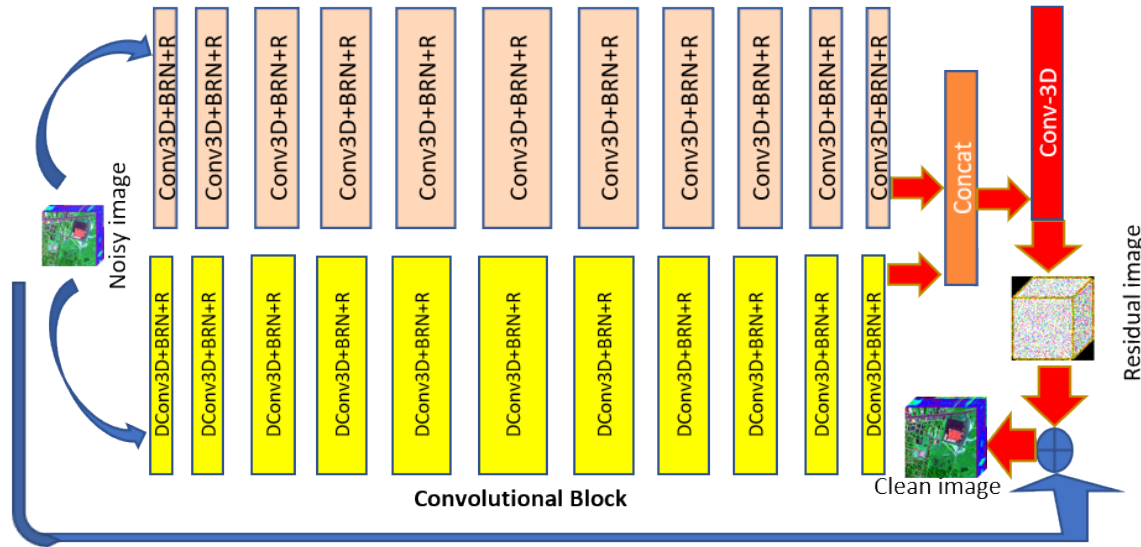
FIGURE 1. The proposed neural network architecture.

filters and decreasing the depth of neural network. Hence, along the two paths, the number of filters is symmetrically increased to 128 and then decreased to 1 after the concatenation layer and the number of serially arranged layers is fixed at 11. At the end junction of the two parallel paths, the concatenation layer doubles the number of features. Hence, one more 3D convolution layer is added at the end to obtain image of the same size that was initially fed into the model. A residual connection at the end with the initial data helps in learning residual image [48]; as indicated in the loss function (Eq. (9)) under section IV.

Combination of residual learning (RL) [48] and batch normalization (BN) [47] is established to obtain good results for image restoration tasks. The popularity of BN is essentially attributed to its ability to mitigate internal covariate shift in data passed along subsequent layers [49]; from one epoch to the next during the training process. As a result, BN is added between convolutional and activation layers. However, it is observed that BN provides biased results when data in input batch are non-independent and identically distributed (non-iid). Hence, in this paper, we used Batch Renormalization (BRN) [50] rather than BN to mitigate the effect of internal co-variate shift and allow usage of non-iid data in our training data preparation. This is especially important in our case where we train two different models for homogeneous and heterogeneous noise levels. Under homogeneous noise settings, noise level among different training samples can be different. Further, under heterogeneous noise settings during training, noise levels differ even among layers within the same sample. All these modifications in the training data violate the iid assumption required to be maintained for a model trained using BN. Under such circumstances, BRN is a more feasible option; which provides fair results even when this restriction is violated. Hence, in this paper, we used BRN rather than BN between convolutional and activation layers along both the paths.
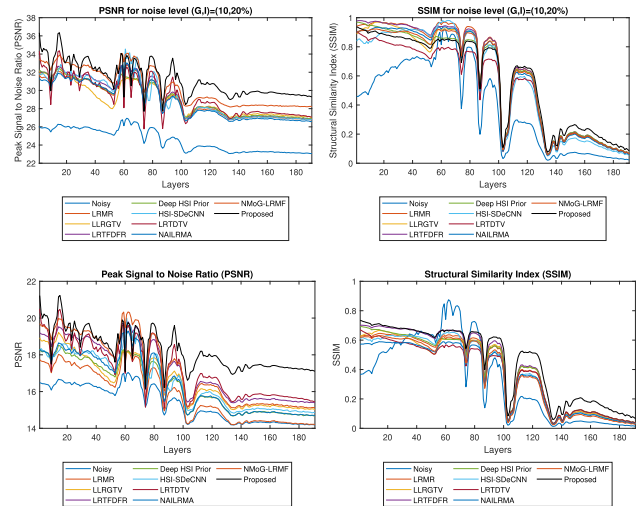

FIGURE 2. Layerwise PSNR and SSIM for WDC dataset for homogeneous $((G, I) = (10 \text{ dB}, 20\%))$ (row 1) and heterogeneous noise (row 2) levels.

Adaptive Moment estimation (ADAM) [51] is used as the optimization technique for training the model. Parameters of the optimizer is set to be $\alpha = 1e - 3$, $\beta_1 = 0.9$, $\beta_2 = 0.999$ and $\epsilon = 1e - 7$. Please note that the term α used here is the learning rate in ADAM and is different from the weighting term α used in the proposed section. Batch-size is set at 30 samples. The model is trained for 200 epochs until convergence.

B. TRAINING DATA

Two separate models are trained. Under homogeneous case, every layer in 3D training sample is corrupted by same noise level and under heterogeneous case, every layer in a training sample is corrupted by a different noise level. For training data preparation, the range of Gaussian noise is varied between [20 dB $\cdots$ 5 dB] and impulse noise is varied

between [5% ⋯ 25%]. Images used for training are obtained from CAVE[1] and Interdisciplinary Computer Vision Laboratory (ICVL) [52] [2] multi-/hyperspectral datasets. Images from CAVE dataset are obtained from Cooled CCD camera (*Apogee Alta U260*) which is a multi-spectral camera acquiring images of size $512 \times 512 \times 31$ in the wavelength range 400 nm to 700 nm in steps of 10 nm. There are in total 32 images in the dataset obtained in artificial lighting conditions in the 16-bit format. Images in ICVL are acquired from *Specim PS Kappa DX4* hyperspectral camera which contain more than 200 images (at the time of preparation of this manuscript). The dataset accommodates images from varying acqusition environments: urban, sub-urban, indoor and plant life. Size of every image is $1392 \times 1300 \times 519$ over wavelength range 400 nm to 1000 nm with increments of 1.25 nm. 15 images from CAVE and 15 images from ICVL are used for training and rest are used for testing. 20% of the training data is used as the validation set. Motivated by the efficient restoration under patch-based training [53], [54], in this work we have extracted 3D patch-size of dimension $40 \times 40 \times 6$. Since non-overlapping patches are used, total size of training data is $12690 (= (512/40 \times 512/40 \times 31/6) \times 15)$ for CAVE dataset and $978310 (= (1390/40 \times 1300/40 \times 519/6) \times 15)$ for ICVL dataset. This gives a total of 991000. Every training sample is normalized in the range $[0 \cdots 1]$ before being fed to the model.

V. EXPERIMENTS AND DISCUSSION

In this section, we illustrate experimental results on three representative artificially corrupted datasets and two real hyperspectral data using eight state-of-the-art techniques and four quantitative metrics. More details are present in the subsequent sections.

A. EXPERIMENTAL SETUP

For quantitative evaluation, peak signal to noise ratio (PSNR) and Structural Similarity Index (SSIM) [55] are used as spatial metrics calculated and averaged over all bands of 3D cube. To measure reduction in spectral distortion observed while noise removal, two spectral metrics are employed as well: Mean Spectral Angle Mapper (MSAM) and Erreur Relative Globale Adimensionnelle de Synthèse (ERGAS) [56]. Higher values of PSNR and SSIM and lower values of MSAM and ERGAS denote better reconstruction quality. To show the effectiveness of the proposed denoising solution; six model-based and two learning-based HSI denoising methods are used for visual and quantitative evaluation. Low-Rank Matrix Recovery (LRMR) [6], LLRGTV (local low-rank matrix recovery and global spatial-spectral total variation) [19], LRTF-DFR (Double-Factor-Regularized Low-Rank Tensor Factorization) [24], LRTDTV (Hyperspectral Image Restoration Via Total Variation Regularized Low-Rank Tensor Decomposition) [57], NAILRMA

(Hyperspectral Image Denoising via Noise-Adjusted Iterative Low-Rank Matrix Approximation) [58] and NMoG-LRMF (Denoising Hyperspectral Image With Non-i.i.d. Noise Structure) [5] are model-driven approaches. Deep HSI Prior [27] and HSI-SDeCNN (A Single Model CNN for Hyperspectral Image Denoising) [29] are data-driven techniques.

B. EXPERIMENTS ON SYNTHETIC DATA

Washington DC (WDC) dataset[3] obtained from Hyperspectral Digital Imaging Collection Experiment (HYDICE) sensor [59] is a representative HSI dataset with 191 spectral bands. Similarly, one CAVE dataset ($512 \times 512 \times 31$) and an ICVL dataset obtained in artificial and natural settings respectively are tested in our synthetic experiments. Layerwise PSNR and SSIM for homogeneous $(G, I) = (10$ dB, $20\%)$ and heterogeneous noise levels for WDC dataset are shown in Fig. 2. With exceptional intermittent drops in metric values (PSNR and SSIM), our proposed solution is able to maintain considerable gap with competing methods under both homogeneous and heterogeneous noise scenarios. Further, we can observe that LLRGTV and Deep HSI Prior are second and third best performing techniques. To conduct a more generalised examination; averaged metric results are tabulated in Table 2. For homogeneous noisy cases, results are represented after averaging over all layers for five different noise levels $(G, I) = (20$ dB, $5\%)$, $(15$ dB, $10\%)$, $(12$ dB, $15\%)$, $(10$ dB, $20\%)$ and $(5$ dB, $25\%)$. Table 2 shows PSNR, SSIM, MSAM and ERGAS results. As shown in the table, we can observe that our proposed method yields the best results for most levels of noise. For heterogeneous noisy cases, results are generated after artificially corrupting each layer of 3D cube with separate noise levels uniformly randomly obtained in the range $G = [20$ dB $\cdots 5$ dB$]$ and $I = [5\% \cdots 25\%]$. Metrics results for heterogeneous noise cases are tabulated in Table 3 and 4. Since the layers are corrupted by mixture noise among different layers, values obtained are lesser than those observed in Table 2.

Visual analysis under different techniques are presented in Fig. 3 and Fig. 4 for homogeneous and heterogeneous noise settings respectively for WDC dataset. Zoomed sections in images show the blurring effect in Deep HSI Prior and artefacts prominent in HSI-SDeCNN. LRMR is unable to remove grainy texture from noisy data. For all the three datasets and eight comparing techniques; our proposed method is able to remove noise while retaining details in the image. This is attributed to the TV term being augmented to the original loss function during training. TV is known for its edge preserving capability at high-frequency regions in the image while removing noise from piece-wise constant regions.

To measure the reconstruction accuracy of the proposed denoising model against competing methods, we have plotted the spectral signature of all techniques in Fig. 5 for both homogeneous and heterogeneous noise cases corresponding to Fig. 3 and 4 respectively. Prior to plotting the graphs,

---

[1] https://www.cs.columbia.edu/CAVE/databases/multispectral/
[2] http://icvl.cs.bgu.ac.il/hyperspectral/

[3] http://lesun.weebly.com/hyperspectral-data-set.html

TABLE 2. Mean PSNR, mean SSIM, mean SAM and ERGAS (All layers corrupted by same level of noise).

Gaussian (dB)	Impulse (%)	Methods	WDC				CAVE				ICVL			
			PSNR	SSIM	MSAM	ERGAS	PSNR	SSIM	MSAM	ERGAS	PSNR	SSIM	MSAM	ERGAS
20	5	Noisy	33.81	0.825	-	-	30.45	0.810	-	-	30.79	0.78	-	-
		LRMR	35.41	0.899	5.13	13.95	32.69	0.860	12.52	23.00	36.05	0.883	6.08	10.17
		LLRGTV	36.86	0.854	12.15	37.16	31.11	0.829	13.26	21.22	37.79	0.903	10.51	28.81
		LRTF-DFR	43.49	0.944	3.95	4.04	34.71	0.900	5.84	5.27	38.07	0.914	2.83	7.03
		Deep HSI Prior	37.72	0.910	5.72	13.25	34.96	0.877	6.18	14.43	38.41	0.929	4.89	9.77
		HSI-SDeCNN	40.12	0.912	9.16	28.21	36.87	0.950	4.82	5.13	40.77	0.940	6.05	11.11
		LRTDTV	42.42	0.891	5.36	20.82	36.03	0.927	5.79	5.66	40.91	0.937	4.79	8.42
		NAILRMA	39.33	0.838	4.39	12.44	35.91	0.938	5.72	5.89	40.76	0.932	5.91	8.43
		NMoG-LRMF	40.77	0.921	4.10	10.89	**39.32**	0.942	5.41	5.47	**41.38**	0.943	4.48	7.39
		Proposed	**43.86**	0.935	4.05	5.94	37.23	**0.961**	4.56	4.71	41.33	**0.951**	2.93	6.39
15	10	Noisy	28.86	0.706	-	-	29.70	0.746	-	-	29.44	0.766	-	-
		LRMR	34.04	0.866	9.06	28.00	33.77	0.866	9.05	15.36	34.41	0.854	7.03	12.61
		LLRGTV	33.77	0.850	9.34	26.51	33.50	0.849	14.01	24.21	34.75	0.862	6.06	12.09
		LRTF-DFR	34.42	0.863	7.98	2.83	36.08	0.873	6.07	10.95	35.09	0.862	8.71	4.05
		Deep HSI Prior	33.89	0.832	8.65	22.38	34.08	0.753	8.64	18.69	34.10	0.785	4.76	11.03
		HSI-SDeCNN	34.92	0.855	12.62	41.00	37.63	0.891	6.07	8.31	35.60	0.860	7.56	15.23
		LRTDTV	35.67	0.849	10.39	30.21	34.72	0.901	6.72	8.99	35.91	0.849	8.21	13.81
		NAILRMA	34.23	0.836	13.22	28.82	36.44	0.899	6.32	8.56	34.58	0.834	8.52	12.70
		NMoG-LRMF	35.77	0.851	10.17	20.21	37.78	0.911	6.92	8.48	34.92	0.872	8.51	11.92
		Proposed	**36.34**	**0.890**	7.92	3.99	**38.70**	**0.919**	5.78	8.34	**37.96**	**0.880**	4.41	**4.01**
12	15	Noisy	27.69	0.687	-	-	28.51	0.616	-	-	28.27	0.615	-	-
		LRMR	30.81	0.707	11.73	38.04	32.04	0.816	11.71	19.43	31.03	0.778	8.27	17.26
		LLRGTV	30.72	0.690	11.22	34.34	31.84	0.806	11.19	15.69	31.42	0.794	9.65	24.78
		LRTF-DFR	31.41	0.711	**7.71**	5.55	32.44	0.828	6.14	10.69	32.21	0.818	4.90	**4.53**
		Deep HSI Prior	**33.89**	0.732	8.65	22.38	33.17	0.714	6.99	16.33	31.68	0.747	6.98	10.64
		HSI-SDeCNN	31.53	0.679	16.33	57.43	33.28	0.831	8.49	11.20	32.25	**0.829**	8.80	18.03
		LRTDTV	33.72	**0.744**	15.87	39.72	34.42	0.857	8.44	12.89	32.79	0.811	9.32	16.03
		NAILRMA	31.79	0.713	16.33	32.91	32.67	0.822	9.01	10.89	32.43	0.801	9.98	18.92
		NMoG-LRMF	32.41	0.734	12.92	27.65	33.22	0.853	8.32	**9.31**	32.78	0.805	7.43	12.74
		Proposed	32.47	0.743	7.84	**5.49**	**34.95**	**0.872**	5.90	9.61	**33.22**	0.823	**4.68**	4.99
10	20	Noisy	24.53	0.507	-	-	25.33	0.514	-	-	26.10	0.507	-	-
		LRMR	29.01	0.669	16.33	59.41	28.84	0.746	12.90	22.70	28.57	0.706	13.97	24.55
		LLRGTV	28.97	0.651	12.67	40.61	26.98	0.746	11.58	16.38	28.99	0.732	12.38	29.78
		LRTF-DFR	29.30	0.673	**8.34**	6.63	30.51	0.787	6.71	11.54	30.14	0.778	12.45	6.08
		Deep HSI Prior	29.11	0.642	9.89	13.96	29.57	0.673	7.85	17.39	29.83	0.714	7.59	10.40
		HSI-SDeCNN	29.24	0.624	18.74	70.02	31.01	0.781	9.18	12.26	30.03	0.739	7.55	19.57
		LRTDTV	29.32	0.672	16.78	59.81	**32.44**	0.782	9.10	14.67	30.11	0.771	7.39	17.03
		NAILRMA	28.91	0.644	18.92	62.34	30.91	0.724	10.92	12.58	29.91	0.764	8.72	16.02
		NMoG-LRMF	30.44	0.641	10.23	43.44	30.21	0.756	7.09	10.01	30.11	0.741	7.56	12.58
		Proposed	**30.85**	**0.692**	8.59	**6.24**	32.26	**0.796**	6.68	10.96	**30.23**	**0.784**	7.36	7.26
5	25	Noisy	23.60	0.436	-	-	23.86	0.478	-	-	23.81	0.487	-	-
		LRMR	27.51	0.532	19.83	78.87	28.37	0.676	15.66	29.89	27.42	0.625	10.38	25.44
		LLRGTV	27.48	0.514	15.12	50.41	28.67	0.702	13.84	22.04	27.95	0.681	10.59	29.78
		LRTF-DFR	27.54	0.540	9.95	8.68	28.73	0.736	8.29	14.03	28.37	0.736	9.79	**6.08**
		Deep HSI Prior	27.38	0.508	10.37	28.33	28.07	0.633	6.86	13.94	28.14	0.677	10.49	10.40
		HSI-SDeCNN	27.31	0.510	18.21	67.08	28.97	0.747	7.51	**8.26**	28.20	0.733	11.61	19.57
		LRTDTV	27.33	0.611	12.59	60.92	28.72	0.738	7.99	9.21	29.03	0.746	10.23	20.98
		NAILRMA	27.22	0.621	13.92	65.92	28.65	0.730	8.35	10.82	29.21	0.723	11.34	29.82
		NMoG-LRMF	28.87	0.650	10.23	48.93	28.73	0.758	7.83	8.67	30.03	0.750	10.92	13.93
		Proposed	**31.35**	**0.677**	8.81	**7.02**	**30.01**	**0.764**	7.33	13.53	**30.12**	**0.769**	8.34	5.26

TABLE 3. Mean PSNR and SSIM for heterogeneous noise levels.

	WDC		CAVE		ICVL	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
Noisy	15.34	0.33	17.17	0.43	18.40	0.59
LRMR	15.83	0.38	18.08	0.54	20.70	0.60
LLRGTV	16.58	0.38	19.90	0.61	21.40	0.65
LRTF-FR	17.20	0.41	21.57	0.70	22.94	0.76
HSI Prior	15.30	0.39	23.13	0.76	22.29	0.72
HSI-SDeCNN	15.55	0.37	24.23	0.79	22.56	0.71
LRTDTV	17.21	0.36	23.72	0.67	22.87	0.69
NAILRMA	16.45	0.35	22.98	0.61	22.45	0.61
NMoG-LRMF	17.23	0.40	23.87	0.64	22.89	0.67
Proposed	**17.31**	**0.45**	**24.50**	**0.82**	**23.08**	**0.79**

TABLE 4. Mean SAM and ERGAS for heterogeneous noise levels.

	WDC		CAVE		ICVL	
	MSAM	ERGAS	MSAM	ERGAS	MSAM	ERGAS
LRMR	20.24	78.01	21.76	31.82	13.97	40.31
LLRGTV	18.14	57.12	18.31	25.56	12.38	31.57
LRTF-FR	8.36	12.56	11.12	16.90	12.45	12.68
HSI Prior	13.01	32.49	**7.90**	14.08	5.94	10.26
HSI-SDeCNN	18.71	68.39	11.12	8.52	7.55	14.66
LRTDTV	14.21	30.22	10.79	**7.32**	7.98	15.87
NAILRMA	16.65	33.76	11.78	8.21	8.34	16.54
NMoG-LRMF	12.78	21.77	10.65	7.91	7.43	14.51
Proposed	**8.19**	**12.37**	11.03	8.46	**5.78**	**10.17**

all layers are normalized in the range [0 · · · 255]. Fig. 5 shows the spectral signature for pixel location (160, 140) at noise level $(G, I) = (10$ dB , $20\%)$. For heterogeneous noise settings, plots are shown in the same figure for pixel location (56, 110). From these figures, it is evident that

the proposed technique reconstructs pixels in the accurate radiometric range compared to that of the original spectral signature of the pixel at the specified locations. The results of LLRGTV and LRTFDFR are also appreciable with the notable exception of LRMR. From the graphs for LRTDTV, it is evident that the reconstruction suppresses local changes in pixel intensities to create a smoother curve.
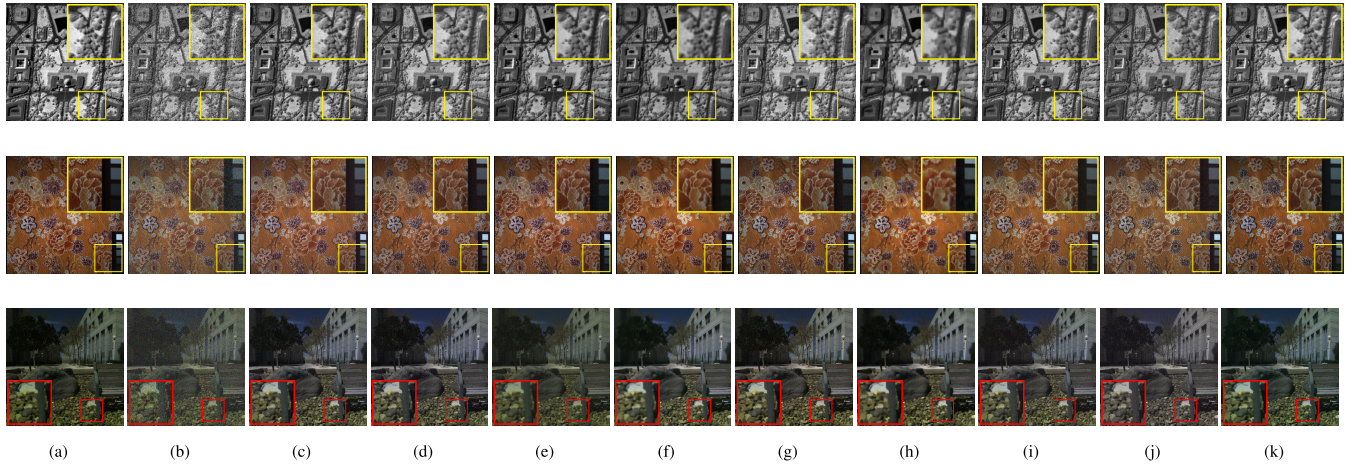
FIGURE 3. Denoising results for Washington DC (row 1), CAVE (row 2) and ICVL (row 3) datasets for noise levels (G , I) = (10 dB , 20%), (G , I) = (2 dB , 30%) and (G , I) = (5 dB , 25%) respectively; with same noise levels added to all the layers. (a) Original, (b) Noisy, (c) LRTV, (d) LLRTGTV, (e) LRTF-DFR, (f) Deep HSI Prior, (g) HSI-SDeCNN, (h) LRTDTV, (i) NAILRMA, (j) NMoG-LRMF, and (k) Proposed.

TABLE 5. PSNR and SSIM for different levels of Gaussian+stripe noise for CAVE dataset.

Noise levels (SNR,%)	Gaussian (**G** (in dB)), Stripe (**S** (in %)) noise		
	20 dB, 20%	15 dB, 30%	10 dB, 40%
Peak Signal to Noise Ratio (PSNR)			
Noisy	32.46	29.85	25.48
LRTV	39.62	38.24	36.59
LLRGTV	42.38	40.41	39.89
LRTFDFR	38.70	37.07	36.56
Deep HSI Prior	42.70	40.82	38.96
HSI-SDeCNN	35.29	34.34	32.15
LRTDTV	38.40	37.62	37.90
NAILRMA	35.61	34.54	32.65
NMoG-RPCA	39.41	37.93	35.67
Proposed	**43.81**	**41.40**	**40.33**
Structural Similarity Index (SSIM)			
Noisy	0.73	0.64	0.58
LRTV	0.86	**0.85**	**0.84**
LLRGTV	0.78	0.71	0.64
LRTFDFR	0.85	0.81	0.78
Deep HSI Prior	0.80	0.78	0.71
HSI-SDeCNN	0.77	0.67	0.62
LRTDTV	0.83	0.80	0.72
NAILRMA	0.79	0.71	0.68
NMoG-RPCA	0.84	0.78	0.75
Proposed	**0.89**	**0.85**	0.83

However, this representation does not reflect diversity in the spectral signature of the original image. The deep learning technique Deep HSI Prior yields better results than HSI-SDeCNN for layers after 120. We obtain the same observations for the proposed model trained over heterogeneous noise levels in Fig. 5. The results of LLRGTV is nearly accurate in this case.

1) STRIPE NOISE REMOVAL
Stripe noise is usually encountered for remotely sensed hyperspectral images obtained using wiskbroom and pushbroom sensors. It also arises during inconsistent response from image detectors and when sensors go out of the radiometric

range [60]. In order to simulate the effect of stripe noise (as faced in real HSI data), we made use of the method proposed in [60], [61]. The data is corrupted by a combination of Gaussian+stripe noise. Image from CAVE dataset [62] is used in our simulation experiments. The data is corrupted with three different levels of Gaussian noise (20 dB, 15 dB and 10 dB) followed by three different levels of stripe noise (10%, 15% and 20%); respectively. Percentage levels in stripe noise indicate the percentage of total number of layers randomly selected as potential candidates for stripe noise. From these randomly selected layers, 20−40 columns are randomly selected to be added with stripe noise. For those columns, the strength of the signal is either increased or decreased with the mean intensity level of the data. Results for CAVE dataset at noise level (10 dB, 20%) are shown in Fig. 6. Results of quantitative metrics are tabulated in Table 5.
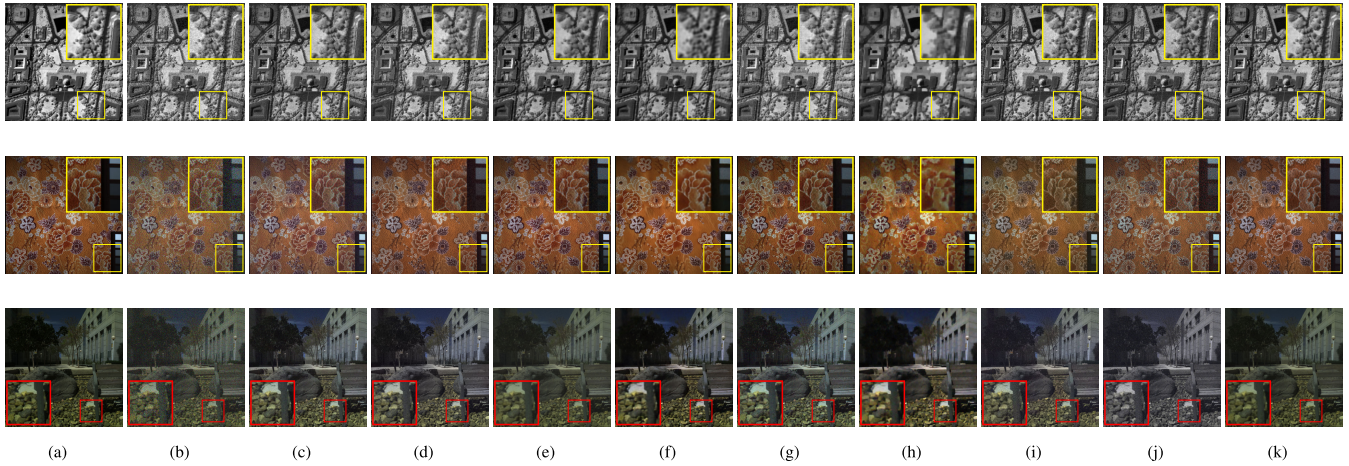
C. EXPERIMENTS ON REAL DATA
For experiments over real data; two datasets are used. First is the *Urban* dataset[4] (Fig. 7). It is obtained in the wavelength range 440 nm to 2500 nm with 210 spectral bands and 307×307 pixels in the spatial resolution covering a ground area of 2 *m* × 2 *m*. Our second dataset is *Salinas* dataset[5] obtained from Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) sensor covering a ground area of $3.7m^2$ with a spatial resolution of 512×217 pixels for 224 spectral bands.

Three zoomed sections of *Urban* dataset are visualized in Fig. 7 depicting the reconstructed image quality on low- and high-textured regions. On the one hand, high level of artefacts can be seen in HSI-SDeCNN while LRMR is unable to remove horizontal stripes often encountered when sensors go out of the radiometric range. On the other hand, HSIPrior

---

[4]https://rslab.ut.ac.ir/data
[5]http://www.ehu.eus/ccwintco/index.php/Hyperspectral-Remote-Sensing-Scenes

FIGURE 4. Denoising results for Washington DC (row 1), CAVE (row 2) and ICVL (row 3) datasets for heterogeneous noise levels. (a) Original, (b) Noisy, (c) LRTV, (d) LLRTGTV, (e) LRTF-DFR, (f) Deep HSI Prior, (g) HSI-SDeCNN, (h) LRTDTV, (i) NAILRMA, (j) NMoG-LRMF, (k) Proposed method.
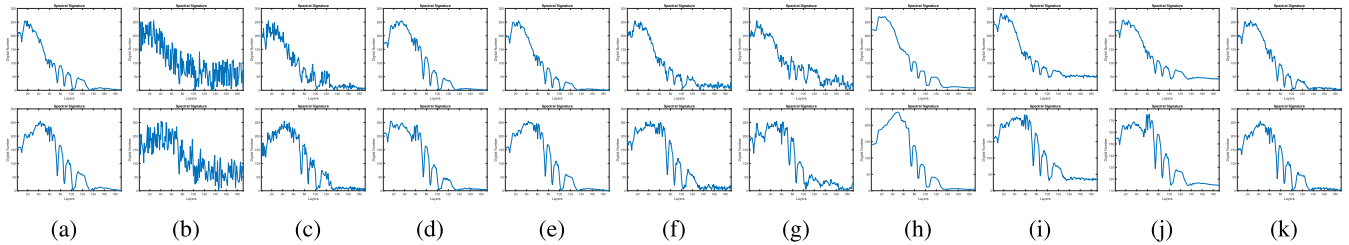
FIGURE 5. Row 1: Spectral Signature for WDC dataset at noise level $(G, I) = (10\ dB, 20\%)$ at pixel location (160, 140) for different methods. Row 2: Spectral Signature for WDC dataset for heterogeneous noise levels at pixel location (56, 110) for different methods.(a), Ground truth, (b) Noisy, (c) LRMR, (d) LLRGTV, (e) LRTF-DFR, (f) Deep HSI Prior, (g) HSI-SDeCNN, (h) LRTDTV, (i) NAILRMA, (j) NMoG-LRMF and (k) Proposed.
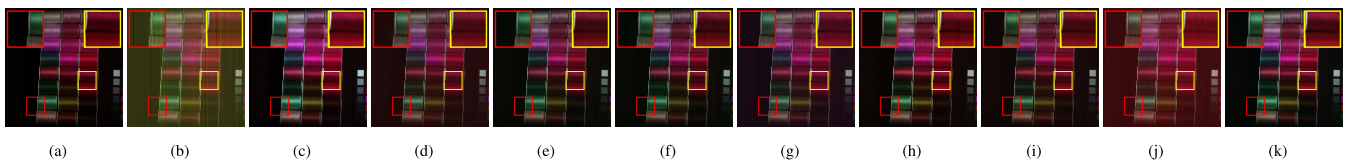
FIGURE 6. Restoration results of experiments conducted on synthetically corrupted (Gaussian+stripe noise) CAVE dataset. (a) Clean, (b) Noisy, (c) LRTV, (d) LLRGTV, (e) LRTFFR, (f) Deep HSI Prior, (g) HSI-SDeCNN, (h) LRTDTV, (i) NAILRMA, (j) NMoG-LRMF and (k) Proposed method.

TABLE 6. Classification results using SVM for different techniques over *Salinas* dataset.

	Overall Accuracy (%)	Kappa Coefficient
Noisy	74.007	0.7080
LRMR	84.925	0.8314
LLRGTV	95.194	0.9376
LRTF-DFR	94.067	0.9250
Deep HSI Prior	80.115	0.7744
HSI-SDeCNN	91.183	0.8946
LRTDTV	89.24	0.8799
NAILRMA	85.45	0.8371
NMoG-LRMF	86.70	0.8479
Proposed	**96.472**	**0.9480**

introduces false colour (not present in the original data) into the resultant denoised image. However, our proposed technique is capable of removing noise while still keeping the details intact.

Results on *Salinas* dataset are visualized in Fig. 8. Insufficient denoising capability is quite evident in data restored using LRMR. Deep HSI Prior removes noise but blur is introduced in the end result. LRTF-DFR and HSI-SDeCNN recover the image quite well but colour reproduction is not the same as present in the original data. Additionally, artefacts (visible as pixelated image in the zoomed image sections) in visual result denoised using HSI-SDeCNN suggest limited denoising performance of the technique. Finally, the results produced by our proposed method is capable of restoring the data with sufficiently good visual quality without introducing any post-processing artefacts.

1) RESULTS ON CLASSIFICATION ACCURACY
Since *Urban* dataset is essentially designed as a benchmark dataset for spectral unmixing techniques, it cannot be used to
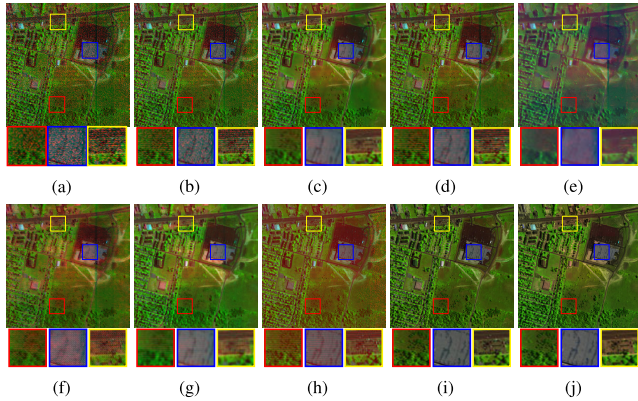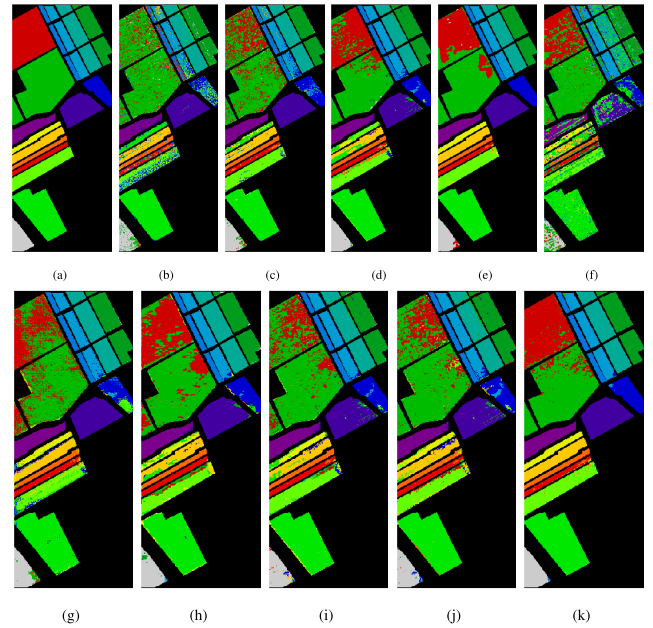
FIGURE 7. Denoising results for urban dataset (a) Noisy, (b) LRMR, (c) LLRTGTV, (d) LRTF-DFR, (e) Deep HSI Prior, (f) HSI-SDeCNN, (g) LRTDTV, (h) NAILRMA, (i) NMoG-LRMF and (j) Proposed. Pseudo-colour image is obtained by the concatenation of bands (208, 70, 20).
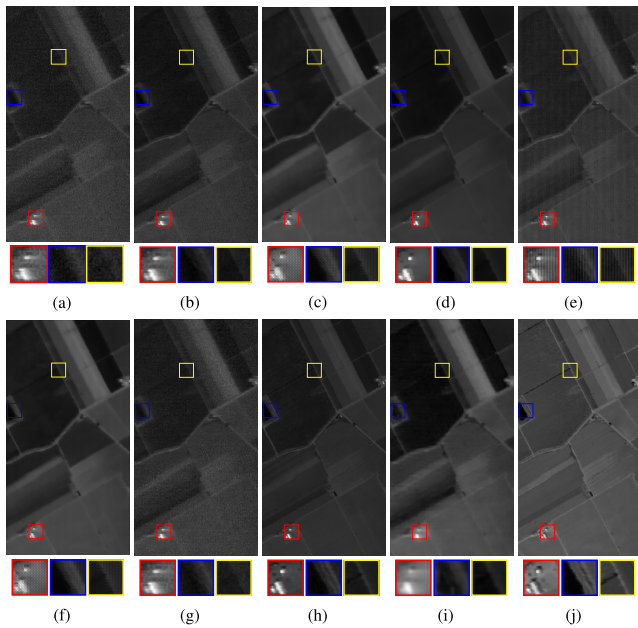
FIGURE 8. Denoising results for salinas dataset (a) Noisy, (b) LRMR, (c) LLRTGTV, (d) LRTF-DFR, (e) Deep HSI Prior, (f) HSI-SDeCNN, (g) LRTDTV, (h) NAILRMA, (i) NMoG-LRMF and (j) Proposed.

FIGURE 9. Classification maps after the application of different denoising methods on Salinas scene dataset. (a), Ground truth, (b) Noisy, (c) LRMR, (d) LLRGTV, (e) LRTF-DFR, (f) Deep HSI Prior, (g) HSI-SDeCNN, (h) LRTDTV, (i) NAILRMA, (j) NMoG-LRMF and (k) Proposed.

**TABLE 7.** Results on running time (in seconds) under different methods.

Methods	Datasets			
	WDC	CAVE	ICVL	Salinas
LRMR	166.83	240.81	241.22	305.46
LLRGTV	135.69	96.26	99.87	585.15
LRTF-DFR	**19.19**	50.33	51.99	**30.74**
Deep HSI Prior	1322.66	1342.44	1348.42	2491.75
HSI-SDeCNN	29.36	47.96	46.88	129.12
LRTDTV	149.13	101.69	104.83	116.55
NAILRMA	89.94	45.44	58.55	266.44
NMoG-LRMF	118.46	222.04	215.80	298.22
Proposed	21.60	**10.51**	**11.46**	44.93

test the classification accuracy of any classification technique on account of non-availability of ground truth classification map (CM). Hence, in this section, we use only *Salinas* dataset to obtain classification results (shown in Fig. 9) over denoised images of different techniques. Support Vector Machine (SVM) is used as the classification algorithm under multi-class label scheme. Since noisy data is present mostly in the end layers of real HSI data, we have used only first 20 layers of the dataset. We have trained the classifier over 20% of labels from each class. Rest of the data is used for testing. Ground truth CM is shown in Fig 9a. With the decrease in restored spatial resolution under Deep HSI prior, we can observe that the former performs worst as far as the classification accuracy is concerned. The presence of artefacts in results obtained by HSI-SDeCNN also affects

its classification accuracy (shown in the ground truth CM highlighted in red colour in top left corner). Overall, from the predicted class labels, we can observe that the classifier trained and tested on the denoising results of the proposed technique is able to classify most of the class labels to their correct regions. It is important to mention that results are obtained using the model trained under heterogeneous noise levels.

To quantify the classification results, values of overall accuracy (OA) and kappa coefficient (KC) are tabulated in Table 6. The metrics results are in direct agreement with visual results in Fig. 9 with Deep HSI Prior performing worst and LLRGTV and LRTF-DFR giving better results. With its ability to outperform the model-driven methods, our proposed technique yields an OA of 96.472 and KC value of 0.9480; as shown in Table 6.

D. EVALUATION OF RUNNING TIME

All the experiments on model-driven techniques (LRMR, LLRGTV, LRTF-DFR, LRTDTV, NAILRMA and

(a)   (b)

(c)   (d)   (e)

(f)   (g)
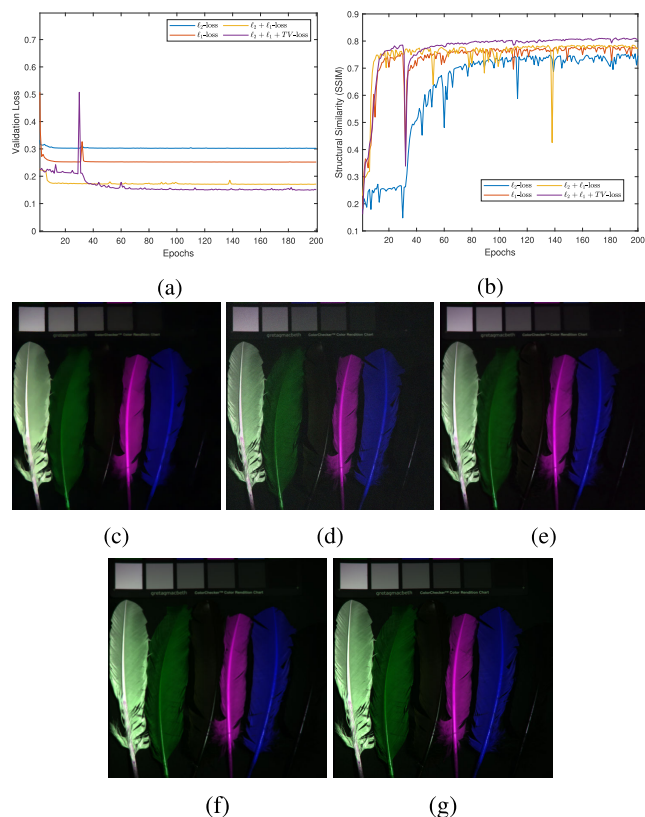
FIGURE 10. (a) Validation loss under different loss criteria, (b) PSNR observed during training over validation dataset and visual results under (c-f) ℓ_2, ℓ_1, $\ell_2 + \ell_1$, $\ell_2 + \ell_1 + TV$ −loss, (g) Ground truth.

NMoG-LRMF) are conducted offline on MATLAB 2019b environment. Training and testing of state-of-the-art and the proposed learning-based technique is performed using GPU Nvidia 16GB Tesla p100 and RAM of 16GB. The total time required for training was around 35 hours for each model. Results of running time over the test datasets are shown in Table 7. Our proposed method was more efficient for CAVE and ICVL dataset. However, the efficiency of LRTF-DFR was better for WDC and *Salinas* dataset. This can be considered an exception when we take into account the overall visual and metric results of the proposed model against competing methods. At the other extreme, Deep HSI Prior is least efficient in running time because of the denoising mechanism used in the approach; which required around 4000 iterations before any useful result could be obtained.

E. ABLATION STUDY

To study the effect of the constituent terms in the loss function used for training the proposed neural network, values of α (Eq. (9)) and λ (Eq. (11)) are carefully tuned to obtain the best results; both visually and in terms of the quality assessment metrics. From Fig. 10a, we can observe that network trained using ℓ_1−loss (setting $\alpha = 0$ in Eq. (9)) shows better reduction in loss over the course of training as compared to ℓ_2−loss ($\alpha = 1$). This is also evident in the visual results; in Fig. 10c

using ℓ_2−loss, estimation is penalised merely based on sum-of-squared pixel level differences generating blurry results and Fig. 10d showing results of ℓ_1−loss trained network where we can observe the effect of insufficient denoising. Combination of $\ell_2 + \ell_1$−loss does provide training (Fig. 10a) and visual (Fig. 10e) benefits but is unable to maintain high-textured details which otherwise is retained by training the network using proposed Bayesian motivated loss function (see Fig. 10a and Fig. 10g). Advantage obtained in terms of SSIM over validation dataset (see Fig. 10b) also substantiate the rationale behind usage of combined $\ell_2 + \ell_1 + TV$ loss providing better metric results during the training period.

VI. CONCLUSION

In this paper, we have designed a discriminative learning framework with a Bayesian viewpoint towards using a loss function. More formally, we have theoretically formulated the data fidelity terms according to Gaussian-impulse corrupted observation in HSI data. Ablation study and experimental results over wide range of noise levels and eight benchmark techniques make the proposed deep learning framework suitable for HSI image restoration on homogeneous and heterogeneous noise scenarios. Incorporating more robust statistical assumptions like Gaussian mixture model and network conditioning can provide further insights towards Bayesian formulation in a learning-based framework.

REFERENCES

[1] L. M. Dale, A. Thewis, C. Boudry, I. Rotar, P. Dardenne, V. Baeten, and J. A. F. Pierna, "Hyperspectral imaging applications in agriculture and agro-food product quality and safety control: A review," *Appl. Spectrosc. Rev.*, vol. 48, no. 2, pp. 142–159, 2013.

[2] G. Lu and B. Fei, "Medical hyperspectral imaging: A review," *J. Biomed. Opt.*, vol. 19, no. 1, 2014, Art. no. 010901.

[3] F. Vasefi, N. MacKinnon, and D. Farkas, "Hyperspectral and multispectral imaging in dermatology," in *Imaging Dermatology*. Amsterdam, The Netherlands: Elsevier, 2016, pp. 187–201.

[4] M. Sandasi, I. Vermaak, W. Chen, and A. Viljoen, "Hyperspectral imaging and chemometric modeling of echinacea—A novel approach in the quality control of herbal medicines," *Molecules*, vol. 19, no. 9, pp. 13104–13121, Aug. 2014.

[5] Y. Chen, X. Cao, Q. Zhao, D. Meng, and Z. Xu, "Denoising hyperspectral image with non-i.i.d. noise structure," *IEEE Trans. Cybern.*, vol. 48, no. 3, pp. 1054–1066, Mar. 2018.

[6] H. Zhang, W. He, L. Zhang, H. Shen, and Q. Yuan, "Hyperspectral image restoration using low-rank matrix recovery," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 8, pp. 4729–4743, Aug. 2014.

[7] P. J. Toivanen, A. Kaarna, J. S. Mielikainen, and M. Laukkanen, "Noise reduction methods for hyperspectral images," *Proc. SPIE Image Signal Process. Remote Sens.*, vol. 4885, pp. 307–313, Dec. 2003.

[8] Y. Qian, Y. Shen, M. Ye, and Q. Wang, "3-D nonlocal means filter with noise estimation for hyperspectral imagery denoising," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, Jul. 2012, pp. 1345–1348.

[9] M. Ye and Y. Qian, "Mixed Poisson-Gaussian noise model based sparse denoising for hyperspectral imagery," in *Proc. 4th Workshop Hyperspectral Image Signal Process. (WHISPERS)*, Jun. 2012, pp. 1–4.

[10] H. Othman and S.-E. Qian, "Noise reduction of hyperspectral imagery using hybrid spatial–spectral derivative-domain wavelet shrinkage," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 2, pp. 397–408, Feb. 2006.

[11] H. Aetesam, K. Poonam, and S. K. Maji, "A mixed-norm fidelity model for hyperspectral image denoising under Gaussian-impulse noise," in *Proc. Int. Conf. Inf. Technol. (ICIT)*, Dec. 2019, pp. 137–142.

[12] H. K. Aggarwal and A. Majumdar, "Generalized synthesis and analysis prior algorithms with application to impulse denoising," in *Proc. Indian Conf. Comput. Vis. Graph. Image Process.*, Dec. 2014, pp. 1–7.

[13] S. Tariyal, H. K. Aggarwal, and A. Majumdar, "Hyperspectral impulse denoising with sparse and low-rank penalties," in *Proc. 7th Workshop Hyperspectral Image Signal Process., Evol. Remote Sens. (WHISPERS)*, Jun. 2015, pp. 1–4.

[14] H. K. Aggarwal and A. Majumdar, "Mixed Gaussian and impulse denoising of hyperspectral images," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Jul. 2015, pp. 429–432.

[15] H. K. Aggarwal and A. Majumdar, "Hyperspectral image denoising using spatio-spectral total variation," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 3, pp. 442–446, Mar. 2016.

[16] H. Aetesam, K. Poonam, and S. K. Maji, "Proximal approach to denoising hyperspectral images under mixed-noise model," *IET Image Process.*, vol. 14, no. 14, pp. 3366–3372, Dec. 2020.

[17] H. Aetesam, S. K. Maji, and J. Boulanger, "A two-phase splitting approach for the removal of Gaussian-impulse noise from hyperspectral images," in *Proc. Int. Conf. Comput. Vis. Image Process.* Singapore: Springer, 2020, pp. 179–190.

[18] W. He, H. Zhang, L. Zhang, and H. Shen, "Total-variation-regularized low-rank matrix factorization for hyperspectral image restoration," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 1, pp. 178–188, Jan. 2016.

[19] W. He, H. Zhang, H. Shen, and L. Zhang, "Hyperspectral image denoising using local low-rank matrix recovery and global spatial–spectral total variation," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 3, pp. 713–729, Mar. 2018.

[20] X. Guo, X. Huang, L. Zhang, and L. Zhang, "Hyperspectral image noise reduction based on rank-1 tensor decomposition," *ISPRS J. Photogramm. Remote Sens.*, vol. 83, no. 9, pp. 50–63, Sep. 2013.

[21] Y.-Q. Zhao and J. Yang, "Hyperspectral image denoising via sparse representation and low-rank constraint," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 1, pp. 296–308, Jan. 2015.

[22] J. Xue, Y. Zhao, W. Liao, and S. G. Kong, "Joint spatial and spectral low-rank regularization for hyperspectral image denoising," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 4, pp. 1940–1958, Apr. 2018.

[23] F. Xu, Y. Chen, C. Peng, Y. Wang, X. Liu, and G. He, "Denoising of hyperspectral image using low-rank matrix factorization," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 7, pp. 1141–1145, Jul. 2017.

[24] Y.-B. Zheng, T.-Z. Huang, X.-L. Zhao, Y. Chen, and W. He, "Double-factor-regularized low-rank tensor factorization for mixed noise removal in hyperspectral image," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 12, pp. 8450–8464, Dec. 2020.

[25] W. Xie and Y. Li, "Hyperspectral imagery denoising by deep learning with trainable nonlinearity function," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 11, pp. 1963–1967, Nov. 2017.

[26] R. Imamura, T. Itasaka, and M. Okuda, "Zero-shot hyperspectral image denoising with separable image prior," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. Workshop (ICCVW)*, Oct. 2019, pp. 1–9.

[27] O. Sidorov and J. Yngve Hardeberg, "Deep hyperspectral prior: Single-image denoising, inpainting, super-resolution," in *Proc. IEEE Int. Conf. Comput. Vis. Workshops*, Dec. 2019, p. 0.

[28] Y. Chang, L. Yan, and W. Liao, "HSI-DeNet: Hyperspectral image restoration via convolutional neural network," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 2, pp. 667–682, Feb. 2018.

[29] A. Maffei, J. M. Haut, M. E. Paoletti, J. Plaza, L. Bruzzone, and A. Plaza, "A single model CNN for hyperspectral image denoising," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 4, pp. 2516–2529, Apr. 2019.

[30] Q. Yuan, Q. Zhang, J. Li, H. Shen, and L. Zhang, "Hyperspectral image denoising employing a spatial–spectral deep residual convolutional neural network," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 2, pp. 1205–1218, Dec. 2018.

[31] Q. Zhang, Q. Yuan, J. Li, X. Liu, H. Shen, and L. Zhang, "Hybrid noise removal in hyperspectral imagery with a spatial–spectral gradient network," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 10, pp. 7317–7329, Dec. 2019.

[32] B. Lin, X. Tao, and J. Lu, "Hyperspectral image denoising via matrix factorization and deep prior regularization," *IEEE Trans. Image Process.*, vol. 29, pp. 565–578, 2020.

[33] J. Jiang, L. Zhang, and J. Yang, "Mixed noise removal by weighted encoding with sparse nonlocal regularization," *IEEE Trans. Image Process.*, vol. 23, no. 6, pp. 2651–2662, Jun. 2014.

[34] T. Huang, W. Dong, X. Xie, G. Shi, and X. Bai, "Mixed noise removal via Laplacian scale mixture modeling and nonlocal low-rank approximation," *IEEE Trans. Image Process.*, vol. 26, no. 7, pp. 3171–3186, Jul. 2017.

[35] K. Gregor and Y. LeCun, "Learning fast approximations of sparse coding," in *Proc. 27th Int. Conf. Int. Conf. Mach. Learn.*, 2010, pp. 399–406.

[36] B. Xin, Y. Wang, W. Gao, and D. Wipf, "Maximal sparsity with deep networks?" 2016, *arXiv:1605.01636*.

[37] Y. Chen, W. Yu, and T. Pock, "On learning optimized reaction diffusion processes for effective image restoration," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 5261–5269.

[38] K. H. Jin, M. T. Mccann, E. Froustey, and M. Unser, "Deep convolutional neural network for inverse problems in imaging," *IEEE Trans. Image Process.*, vol. 26, no. 5, pp. 4509–4522, Jun. 2016.

[39] K. Zhang, W. Zuo, S. Gu, and L. Zhang, "Learning deep CNN denoiser prior for image restoration," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 3929–3938.

[40] K. Zhang, Y. Li, W. Zuo, L. Zhang, L. Van Gool, and R. Timofte, "Plug-and-play image restoration with deep denoiser prior," 2020, *arXiv:2008.13751*.

[41] J.-F. Cai, R. H. Chan, and M. Nikolova, "Two-phase approach for deblurring images corrupted by impulse plus Gaussian noise," *Inverse Problems Imag.*, vol. 2, no. 2, p. 187, 2008.

[42] P. Rodriguez, R. Rojas, and B. Wohlberg, "MIxed Gaussian-impulse noise image restoration via total variation," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Mar. 2012, pp. 1077–1080.

[43] J. Jiang, Z. Wang, C. Chen, and T. Lu, "L1-L1 norms for face super-resolution with mixed Gaussian-impulse noise," in *Proc. IEEE Int. Conf. Acoust., speech signal Process. (ICASSP)*, Mar. 2016, pp. 2089–2093.

[44] A. Mahendran and A. Vedaldi, "Understanding deep image representations by inverting them," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 5188–5196.

[45] F. Yu and V. Koltun, "Multi-scale context aggregation by dilated convolutions," 2015, *arXiv:1511.07122*.

[46] X.-J. Mao, C. Shen, and Y.-B. Yang, "Image restoration using convolutional auto-encoders with symmetric skip connections," 2016, *arXiv:1606.08921*.

[47] P. Liu and R. Fang, "Wide inference network for image denoising via learning pixel-distribution prior," 2017, *arXiv:1707.05414*.

[48] K. Zhang, W. Zuo, Y. Chen, D. Meng, and L. Zhang, "Beyond a Gaussian denoiser: Residual learning of deep CNN for image denoising," *IEEE Trans. Image Process.*, vol. 26, no. 7, pp. 3142–3155, Jul. 2017.

[49] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," 2015, *arXiv:1502.03167*.

[50] S. Ioffe, "Batch renormalization: Towards reducing minibatch dependence in batch-normalized models," in *Proc. Adv. Neural Inf. Process. Syst.*, 2017, pp. 1945–1953.

[51] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*.

[52] B. Arad and O. Ben-Shahar, "Sparse recovery of hyperspectral signal from natural RGB images," in *Proc. Eur. Conf. Comput. Vis.* Heidelberg, Germany: Springer, 2016, pp. 19–34.

[53] D. Zoran and Y. Weiss, "From learning models of natural image patches to whole image restoration," in *Proc. Int. Conf. Comput. Vis.*, Nov. 2011, pp. 479–486.

[54] V. Papyan and M. Elad, "Multi-scale patch-based image restoration," *IEEE Trans. Image Process.*, vol. 25, no. 1, pp. 249–261, Jan. 2016.

[55] 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, Apr. 2004.

[56] L. Wald, *Data Fusion: Definitions Architectures: Fusion Images Different Spatial Resolutions*. Paris, France: Presses des MINES, 2002.

[57] Y. Wang, J. Peng, Q. Zhao, D. Meng, Y. Leung, and X.-L. Zhao, "Hyperspectral image restoration via total variation regularized low-rank tensor decomposition," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 4, pp. 1227–1243, Apr. 2018.

[58] W. He, H. Zhang, L. Zhang, and H. Shen, "Hyperspectral image denoising via noise-adjusted iterative low-rank matrix approximation," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 8, no. 6, pp. 3050–3061, Jun. 2015.

[59] P. A. Mitchell, "Hyperspectral digital imagery collection experiment (HYDICE)," *Proc. SPIE*, vol. 2587, pp. 70–95, Nov. 1995.

[60] X. Liu, X. Lu, H. Shen, Q. Yuan, Y. Jiao, and L. Zhang, "Stripe noise separation and removal in remote sensing images by consideration of the global sparsity and local variational properties," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 5, pp. 3049–3060, May 2016.

[61] 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, Jun. 2015.

[62] F. Yasuma, T. Mitsunaga, D. Iso, and S. K. Nayar, "Generalized assorted pixel camera: Postcapture control of resolution, dynamic range, and spectrum," *IEEE Trans. Image Process.*, vol. 19, no. 9, pp. 2241–2253, Sep. 2010.

HAZIQUE AETESAM received the B.Tech. degree in information technology from Jamia Hamdard, New Delhi, India, in 2013, and the M.Tech. degree in computer science and engineering from the Birla Institute of Technology, Ranchi, India, in 2016. He is currently pursuing the Ph.D. degree in computer science and engineering with the Indian Institute of Technology Patna, India. He is working on the low-level computer vision problems pertaining to hyperspectral imaging and magnetic resonance imaging domains. His current research interests include the amalgamation of model and data-driven methods in low-level vision problems.

SUMAN KUMAR MAJI received the B.Tech. degree in electronics and communication engineering from the West Bengal University of Technology, India, in 2006, the master's degree in Telecommunication networks from the Indian Institute of Technology Kharagpur, India, in 2008, and the Ph.D. degree in computer science from INRIA Bordeaux, France, in 2013.

From 2014 to 2015, he worked as a Research Engineer with the Institute of Hematology, University Paris 7, and INSERM. He is currently working as an Assistant Professor with the Department of Computer Science and Engineering, Indian Institute of Technology Patna, India. He has authored several conferences and journal papers. His research interests include medical imaging, bioinformatics, machine learning, and image processing.

Dr. Maji was a recipient of various research fellowships and awards, such as the European CORDIS Doctoral Fellowship (2010), the Region Aquitaine OPTAD Research Fellowship (2010), the FRM Research Fellowship (2014), and the SERB Early Career Research Award from DST, Government of India (2017).

HUSSEIN YAHIA (Member, IEEE) received the Doctorat de 3eme cycle degree from the University of Paris-Sud, Orsay, France, in 1987, and the Habilitation à Diriger des Recherches (HDR) degree from Paris 13 University, Villetaneuse, France, in 2003.

He is currently the Head of the Geostat (Geometry and Statistics in Acquisition Data) Research Team with the French National Public Research Institute in Computer Science and Applied Mathematics (INRIA), Bordeaux, France. He specializes in the analysis of complex signals and systems using approaches from statistical physics; he is also studying sparse signal representations and filtering with specific application to signals in astrophysics in collaboration with the Astrophysics Laboratory of Bordeaux (analysis of turbulent properties in the interstellar medium). He is involved in many national, European, and international contracts and has been supervising over 15 Ph.D. students. He has authored or coauthored more than 100 publications in international peer-reviewed top journals and conferences.

· · ·