SURE BASED CONVOLUTIONAL NEURAL NETWORKS FOR HYPERSPECTRAL IMAGE DENOISING

Han V. Nguyen, Magnus O. Ulfarsson and Johannes R. Sveinsson

Faculty of Electrical and Computer Engineering
University of Iceland

ABSTRACT

This paper addresses the hyperspectral image (HSI) denoising problem by using Stein's unbiased risk estimate (SURE) based convolutional neural network (CNN). Conventional deep learning denoising approaches often use supervised methods that minimize a mean-squared error (MSE) by training on noisy-clean image pairs. In contrast, our proposed CNN-based denoiser is unsupervised and only makes use of noisy images. The method uses SURE, which is an unbiased estimator of the MSE, that does not require any information about the clean image. Therefore minimization of the SURE loss function can accurately estimate the clean image only from noisy observation. Experimental results on both simulated and real hyperspectral datasets show that our proposed method outperforms competitive HSI denoising methods[1].

Index Terms— Hyperspectral image denoising, unsupervised deep learning, convolutional neural network, Stein's unbiased risk estimate.

1. INTRODUCTION

HSI denoising methods can be roughly categorized into model-based and deep learning-based methods. Model-based methods [1] view denoising as an inverse problem and formulate the denoised HSI as a solution to an optimization problem for which the cost function is usually penalized least squares. The majority of deep learning-based methods are based on supervised learning, which relies on the availability of a clean HSI [2]. Therefore, the performance of those methods depends strongly on image penalties and the amount of clean HSI. However, in remote sensing, the use of supervised learning methods is limited because it is challenging to obtain high-quality noise-free hyperspectral data.

Recently, [3] proposed the deep image prior (DIP) method, which assumes that the structure of the CNN provides a suitable image prior for denoising, super-resolution, and inpainting. The basic idea of DIP for denoising is that CNN offers high impedance to noise and low impedance to an image. As a result, during the optimization process, the solution fits the image before the noise. The paper [4] extended DIP to HSI denoising and got results closed to the results obtained by supervised deep learning methods. However, the main drawback of DIP is its tendency to overfitting if the number of iterations in the optimization process is not carefully chosen. Several regularization techniques have been introduced to overcome overfitting. The combination of DIP and total variation and regularization by denoising [5] (DIP-RED) showed considerable improvements over DIP. The papers [6, 7] proposed to train a deep learning network for denoising RGB images by using SURE [8, 9]. The idea is to avoid overfitting by directly minimizing MSE (the mean squared error between the (unknown) clean image and the estimated image). However, MSE is not computable. Therefore MSE is exchanged for SURE, which is a computable unbiased estimate of the MSE.

In this paper,we develop a SURE loss function using Monte-Carlo SURE [10] for a CNN HSI denoiser. Experiments with simulated and real HSI datasets show that our SURE based CNN method gives better performance in both quantiative metrics and visualization against the competitive methods.

2. SURE BASED CNN DENOISING

Let us vectorize an HSI and denote $\mathbf{y} \in \mathbb{R}^N$ as the noisy measurement of the clean but unknown HSI $\mathbf{x} \in \mathbb{R}^N$, where $N = m \times n \times B$, $m \times n$ is the spatial size, and B is the number of bands. The noisy measurement $\mathbf{y}$ is assumed as follows

$$\mathbf{y} = \mathbf{x} + \mathbf{w},$$

where $\mathbf{w} \sim \mathcal{N}(0, \sigma^2 \mathbf{I}_N)$ is Gaussian noise. Therefore, the denoising problem is to estimate a denoised image from its noisy measurement. Here, we propose a SURE based CNN to estimate $\mathbf{x}$ with $\widehat{\mathbf{x}} = f_\theta(\mathbf{y})$, where $f_\theta(\mathbf{y})$ is the output of the CNN, and θ are the network parameters.

This work was supported in part by the Icelandic Research Fund under Grant 174075-05 and Grant 207233-051, and the University of Iceland Doctoral Fund.

[1]Project codes: https://github.com/hvn2/HSI-Denoising-SURE-CNN

2.1. SURE and Monte-Carlo SURE

SURE $\widehat{R}$ is an unbiased estimate of the MSE of any estimator, $R = \frac{1}{N}\|\mathbf{x} - f_\theta(\mathbf{y})\|_2^2$, i.e., $E[R] = E[\widehat{R}]$, where $E[\cdot]$ is the expectation operator. SURE is given by

$$\widehat{R} = \frac{1}{N}\|\mathbf{y} - f_\theta(\mathbf{y})\|_2^2 - \sigma^2 + \frac{2\sigma^2}{N}\mathrm{div}_{\mathbf{y}}(f_\theta(\mathbf{y})), \quad (1)$$

where $\mathrm{div}_{\mathbf{y}}(f_\theta(\mathbf{y}))$ is the divergence of $f_\theta(\mathbf{y})$ and is defined as

$$\mathrm{div}_{\mathbf{y}}(f_\theta(\mathbf{y})) = \sum_{i=1}^{N}\frac{\partial f_{\theta i}(\mathbf{y})}{\partial y_i}.$$

The explicit computation of $\mathrm{div}_{\mathbf{y}}(f_\theta(\mathbf{y}))$ is possible in some special cases, such as linear and coordinate-wise non-linear estimators. However, for deep learning estimator, the computation of the divergence is hard. Fortunately, [10] introduced the Monte-Carlo SURE that allows the estimation of the divergence term as

$$\mathrm{div}_{\mathbf{y}}(f_\theta(\mathbf{y})) \approx \mathbf{b}^T\left(\frac{f_\theta(\mathbf{y}+\epsilon\mathbf{b}) - f_\theta(\mathbf{y})}{\epsilon}\right), \quad (2)$$

where $\mathbf{b} \in \mathbb{R}^N$ is an i.i.d. Gaussian distribution with zero mean and unit variance, ϵ is a small value that can be chosen in the range of 10^{-5} to 10^{-2}.

In practice, the noise level σ in (1) can be estimated by taking the band-wise mean of $\widehat{\sigma}_i, i = 1, ..., B$, where $\widehat{\sigma}_i$ is calculated using the median absolute deviation estimator in the highest subband (HH) of the wavelet transform as [11]

$$\widehat{\sigma}_i = \frac{\mathrm{median}\left(\left|\mathbf{W}_{(i)}^{HH}\right|\right)}{0.6745}, \quad (3)$$

where $\mathbf{W}_{(i)}^{HH}$ is the 2D wavelet coefficients for the subband HH in the i-th band.

2.2. Network Architecture

To demonstrate the idea of SURE, we use a skip connected CNN [3], which provides a good image prior. The structure of the network is inspired by the network used in [3] with some modifications. We do not use any downsampling and upsampling layers, instead the input image size is kept the same throughout the network. The batch normalization layers in [3] are removed since they do not affect much in our network. The network architecture is shown in Fig. 1. It consists of an input branch and an output branch. The back-bone of input and output branches are K blocks of convolutional and *LeakyReLU* (conv-relu) layers. In each input block, a skipped connection is generated by using conv-relu layers. It is concatenated with the corresponding output block except for the last layer. The output layer with B filters and the *Sigmoid* activation function reconstructs the HSI. The SURE loss function in (1) with the divergence term in (2) is used within the network. We use the Adam optimization method with a learning rate of 0.01. The network is implemented in Tensorflow 2.0.

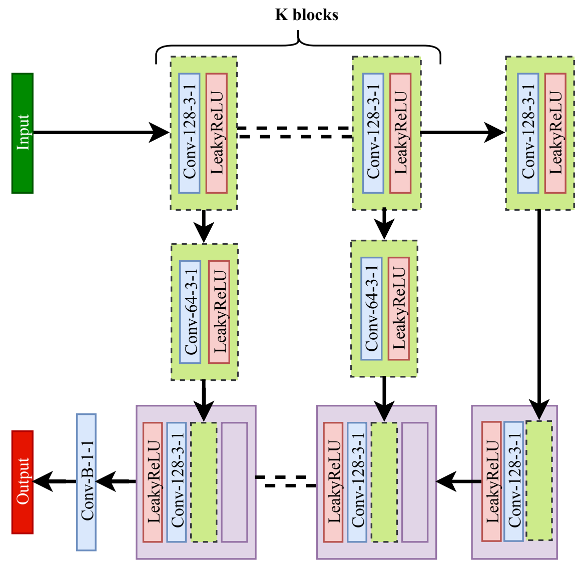

Fig. 1: Network structure. Conv-128-3-1 represents a convolutional layer with 128 filters of kernel size 3 and stride 1. The number of skipped connection blocks is $K = 5$.

3. EXPERIMENTAL RESULTS

In the experiments, we use the Washington DC mall (DC) and the Pavia University (PU) datasets as clean data to generate simulated noisy HSI data, while the Indian Pines (IP) dataset is the real noisy dataset. Two small parts with spatial dimensions of 400×200 are cropped from DC and PU datasets for the experiments. The data are normalized between 0 and 1, band by band, before further processing. The simulated noisy data are created by adding Gaussian noise with zero-mean and standard deviation σ to the DC and PU clean data cubes.

The quantitative metrics are mean SSIM [12], and PSNR in decibels defined as

$$\mathrm{PSNR} = 10\log_{10}\left(\frac{\max^2(\mathbf{x})}{\frac{1}{N}\|\mathbf{x} - \widehat{\mathbf{x}}\|_2^2}\right).$$

Our method, called SURE-CNN, is compared with the following competitive methods: An HSI denoising method using the first-order spectral roughness penalty in the wavelet domain (FORPDN) [12], an HSI restoration technique using sparse and low-rank model (HyRes) [13], and a method based on deep hyperspectral image prior for HSI denoising (DIP) [4]. FORPDN and HyRes use default parameters as recommended in [12] and [13], respectively. DIP and SURE-CNN have the same network structure, but different loss function. DIP uses the fidelity loss (i.e., $\mathcal{L}(\theta)_{\mathrm{DIP}} = \frac{1}{N}\|\mathbf{y} - f_\theta(\mathbf{y})\|_2^2$) while SURE-CNN uses the SURE loss

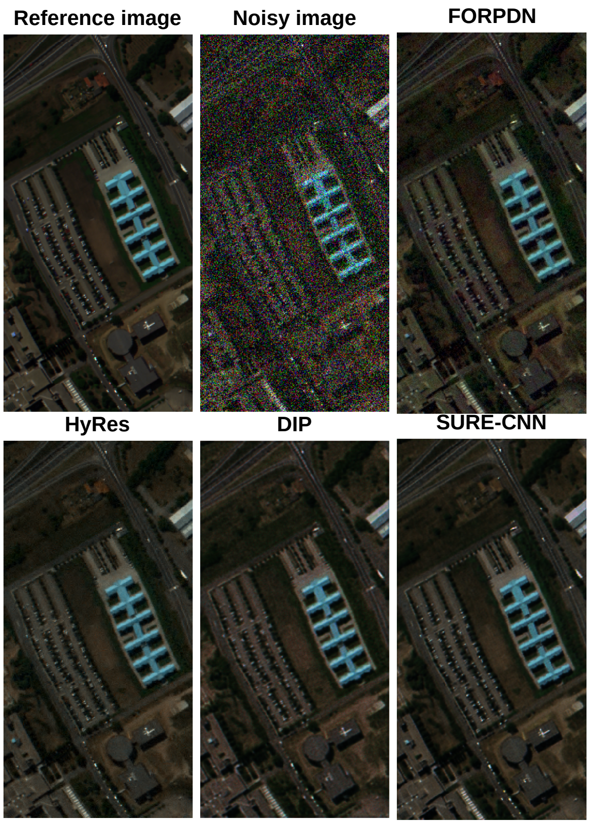

Fig. 2: Denoising results of the simulated PU dataset with $\sigma = 50/255$. The pseudo-color image is created by using bands 56, 33 and 13.

function described in (1) and (2). Fig. 2 shows the denoising results using different methods for the simulated PU dataset and Table 1 shows the results for both PU and DC simulated datasets. It can be seen that SURE-CNN outperforms the competitive methods by higher PSNRs and higher mean SSIMs. All denoising methods provide good quality of denoised images. SURE-CNN gives the denoised image which is closest visualization to the reference image, but the denoised image by FORPDN still has some noise that can be seen by zooming in it. Moreover, SURE loss in SURE-CNN can accurately approximate the true MSE and significantly prevent overfitting over fidelity loss in DIP, as can be seen in Fig. 3, which gives the fidelity loss, the SURE loss and the true MSE of each estimator (i.e., the MSE of the network output and clean image) for denoising PU with $\sigma = 100/255$. For the fidelity loss, the training loss and the true MSE are different by a gap. After 200 iterations, the network tends to overfit where the training loss continues decreasing, but the true MSE begins increasing. In contrast, for the SURE loss, it approximates the true MSE and prevents overfitting.

Finally, SURE-CNN and the competitive methods are applied to denoise the IP dataset. Again, all the denoising methods use the same parameters as for the simulated datasets. But, SURE-CNN and DIP were run for 500 iterations. Denoising results are pictured in Fig. 4. All methods can im-

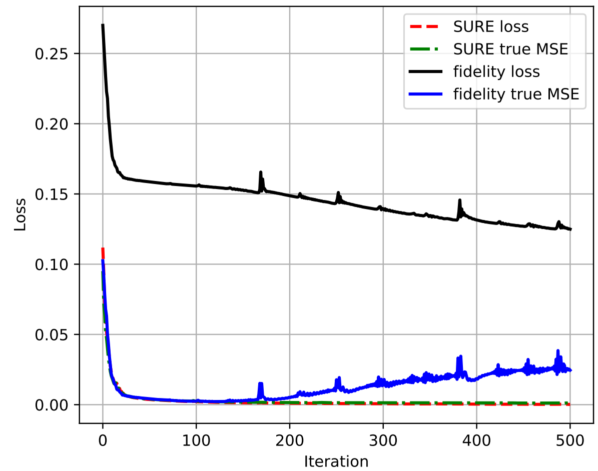

Fig. 3: Training loss using both the fidelity and the SURE loss functions for denoising simulated PU dataset with $\sigma = 100/255$. The SURE loss is a good approximation of the true MSE.

prove the image quality. However, FORPDN does not have good visualization at bands 105 and 220 since it creates water blob-liked noise. HyRes fails to recover band 105. Both DIP and SURE-CNN can significantly remove the noise in the noisy bands such as band 105 and band 220, but DIP seems to be over-bright at band 105. It is worth to evaluate the running time for denoising the IP dataset. FORPDN and HyRes are run under Matlab R2019b on a PC with 8 cores 3.2 GHz CPU and 64 GB RAM. DIP and SURE-CNN are run under Tensorflow 2.0 GPU on a PC with Nvidia Titan X GPU 12 GB. The methods sorted by fastest to slowest are HyRes, FORPDN, DIP, and SURE-CNN with the running times are 1.19, 2.08, 84.32, and 88.90 seconds, respectively.

4. CONCLUSION

We have proposed a new SURE based HSI denoising method. The advantages of SURE are twofold. First, it is an unbiased estimator of the true MSE. Second, using the SURE loss for CNN can significantly prevent overfitting. The experimental results verify our proposed method improve the denoising performance in both simulated noisy and real noisy dataset.

5. REFERENCES

[1] H. Fan, C. Li, Y. Guo, G. Kuang, and J. Ma, “Spatial–spectral total variation regularized low-rank tensor decomposition for hyperspectral image denoising,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 56, no. 10, pp. 6196–6213, 2018.

[2] Q. Yuan, Q. Zhang, J. Li, H. Shen, and L. Zhang, “Hyperspectral image denoising employing a spatial-spectral deep residual convolutional neural network,”

Table 1: Denoising results for the simulated datasets. The table shows PSNR (dB) in first column and mean SSIM in second column for each method. Best results are in boldface type.

Dataset	Noise level	Noisy		FORPDN		HyRes		DIP		SURE-CNN	
PU	$\sigma = 100/255$	8.13	0.0370	26.03	0.5971	28.41	0.7379	26.47	0.6834	**29.62**	**0.8019**
	$\sigma = 50/255$	14.16	0.1286	30.44	0.7989	31.78	0.8549	30.69	0.8460	**33.29**	**0.9048**
	$\sigma = 25/255$	20.17	0.3260	34.34	0.9063	35.35	0.9271	34.48	0.9165	**36.09**	**0.9452**
DC	$\sigma = 100/255$	8.13	0.0498	25.97	0.6678	28.43	0.7923	27.06	0.7650	**30.00**	**0.8594**
	$\sigma = 50/255$	14.15	0.1595	29.54	0.8214	32.24	0.8895	32.01	0.9020	**33.65**	**0.9340**
	$\sigma = 25/255$	20.18	0.3691	32.43	0.8989	36.73	0.9496	36.34	0.9588	**36.80**	**0.9662**

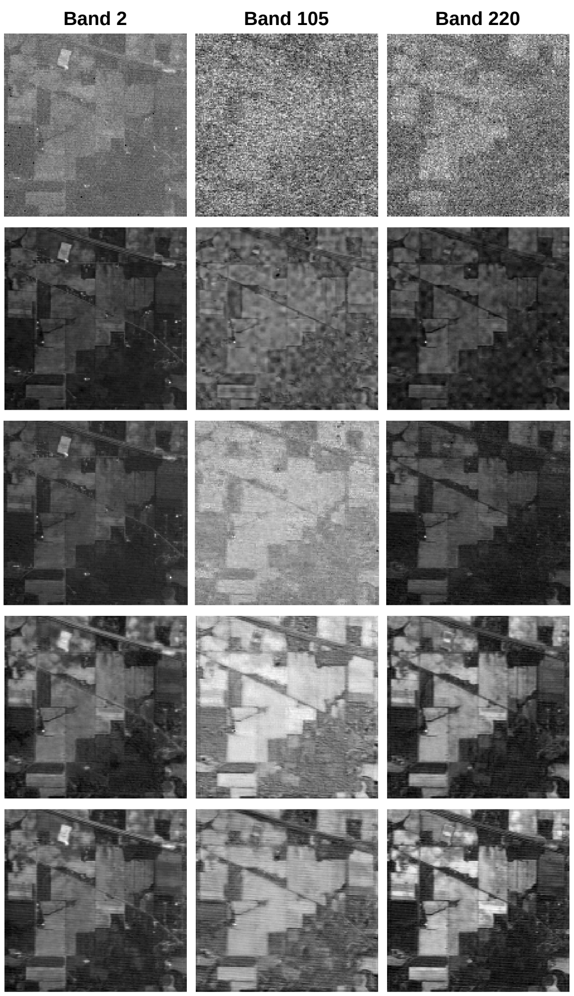

Fig. 4: IP dataset from top to bottom: noisy bands, denoised bands by FORPDN, denoised bands by HyRes, denoised bands by DIP and denoised bands by SURE-CNN.

IEEE Transactions on Geoscience and Remote Sensing, vol. 57, no. 2, pp. 1205–1218, 2018.

[3] D. Ulyanov, A. Vedaldi, and V. Lempitsky, "Deep image prior," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 9446–9454.

[4] O. Sidorov and J. Y. Hardeberg, "Deep hyperspectral prior: Single-image denoising, inpainting, super-resolution," in *The IEEE International Conference on Computer Vision Workshops*, Oct. 2019.

[5] G. Mataev, P. Milanfar, and M. Elad, "DeepRED: Deep image prior powered by RED," in *The IEEE International Conference on Computer Vision Workshops*, Oct. 2019.

[6] C. A. Metzler, A. Mousavi, R. Heckel, and R. G. Baraniuk, "Unsupervised learning with Stein's unbiased risk estimator," in *Proceeding of the International Biomedical and Astronomical Signal Processing Frontiers Workshop*, Feb. 2019, pp. 67–68.

[7] S. Soltanayev and S. Y. Chun, "Training deep learning based denoisers without ground truth data," in *Advances in Neural Information Processing Systems*, 2018, pp. 3257–3267.

[8] C. M. Stein, "Estimation of the mean of a multivariate normal distribution," *The Annals of Statistics*, pp. 1135–1151, 1981.

[9] V. Solo, "A sure-fired way to choose smoothing parameters in ill-conditioned inverse problems," in *Proc. IEEE International Conference on Image Processing (ICIP'96)*, vol. 3, Lausanne, Switzerland, 1996, pp. 89–92.

[10] S. Ramani, T. Blu, and M. Unser, "Monte-Carlo SURE: A black-box optimization of regularization parameters for general denoising algorithms," *IEEE Transactions on Image Processing*, vol. 17, no. 9, pp. 1540–1554, 2008.

[11] S. Mallat, *A wavelet tour of signal processing*. Academic Press, 1998.

[12] B. Rasti, J. R. Sveinsson, M. O. Ulfarsson, and J. A. Benediktsson, "Hyperspectral image denoising using first order spectral roughness penalty in wavelet domain," *IEEE Journal of Selected Topics in Applied Earth*

Observations and Remote Sensing, vol. 7, no. 6, pp. 2458–2467, 2013.

[13] B. Rasti, M. O. Ulfarsson, and P. Ghamisi, "Automatic hyperspectral image restoration using sparse and low-rank modeling," *IEEE Geoscience and Remote Sensing Letters*, vol. 14, no. 12, pp. 2335–2339, 2017.