

Hyperspectral Image Denoising with Realistic Data

Tao Zhang¹ Ying Fu^{1*} Cheng Li²

¹Beijing Institute of Technology ²Huawei Noah's Ark Lab

Abstract

The hyperspectral image (HSI) denoising has been widely utilized to improve HSI qualities. Recently, learning-based HSI denoising methods have shown their effectiveness, but most of them are based on synthetic dataset and lack the generalization capability on real testing HSI. Moreover, there is still no public paired real HSI denoising dataset to learn HSI denoising network and quantitatively evaluate HSI methods. In this paper, we mainly focus on how to produce realistic dataset for learning and evaluating HSI denoising network. On the one hand, we collect a paired real HSI denoising dataset, which consists of short-exposure noisy HSIs and the corresponding long-exposure clean HSIs. On the other hand, we propose an accurate HSI noise model which matches the distribution of real data well and can be employed to synthesize realistic dataset. On the basis of the noise model, we present an approach to calibrate the noise parameters of the given hyperspectral camera. The extensive experimental results show that a network learned with only synthetic data generated by our noise model performs as well as it is learned with paired real data. Our code and data are available at: <https://github.com/ColinTaoZhang/HSIDwRD>.

1. Introduction

Hyperspectral image (HSI) can provide much more spectral information than RGB image, and is beneficial to numerous applications, including remote sensing [6, 35], computer vision [7], medical diagnosis [5, 32] and more. Hyperspectral imaging is to capture spectral information of each spatial position in a scene with massive wavebands, and the commercial hyperspectral cameras often utilize the scanning design [4, 39]. These make its photon counts for each band are much less than that in RGB image, and various noises are easily introduced into the acquisition process. This kind of degradation negatively influences not only the visual appearance of the HSI but also the performance of all downstream HSI applications [14]. Thus, HSI denoising is an essential step in the pipeline of HSI analysis and

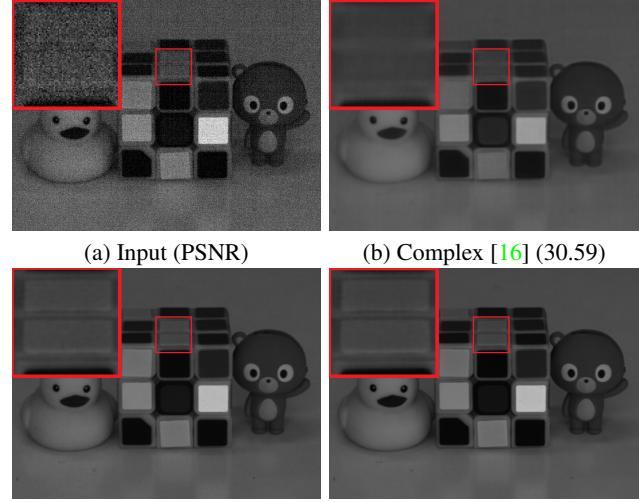


Figure 1. A scene from our collected HSI dataset, and we show the spectral band in 550 nm. (a) The input noisy image; (b) The output of CNN trained with synthetic dataset generated by complex noise model [16]; (c) The output of CNN trained with our collected paired real dataset; (d) The output of CNN trained with synthetic dataset generated by our noise model, which is comparable with (c) the result trained with paired real dataset and obviously outperforms synthetics dataset generated by complex noise model.

processing.

To remove the imaging noise, the well-known model-based HSI denoising methods often iteratively solve an optimization problem with various hand-crafted priors, such as smoothness [50], self-similarity [20, 21] and so on. Nevertheless, the iterative optimization procedure is time-consuming and the hand-crafted priors cannot sufficiently represent the variety of data in the real world. Instead of costly optimization and hand-crafted priors, learning-based methods [9, 19, 29, 45, 51] automatically learn the mapping from noisy HSI to clean HSI with convolutional neural network (CNN). However, existing learning-based methods generally rely on training dataset synthesized with simple Gaussian noise model or complex noise models [16, 45]. Promising results on synthetic data notwithstanding, these methods still cannot well work and evaluate on the real data, due to lacking realistic HSI data.

*Corresponding author: fuying@bit.edu.cn

There are two ways to solve this problem. One is to capture paired real data for HSI denoising network learning and evaluation, like those for RGB image denoising [1, 12, 13, 26, 38]. But collecting abundant high-quality real data for learning HSI denoising network is obviously expensive and requires a large amount of labor. The other is to generate realistic paired data. This is convenient and inexpensive, but it lies in how accurate the noise formulation model of real HSI is. The heteroscedastic Gaussian noise model [2] approximates the noise occurred in HSI better than the commonly-used homoscedastic one. Nevertheless, it cannot delineate the full picture of sensor noise in the HSI. Actually, due to the physical characteristics of hyperspectral camera, e.g., scanning the scene in spatial or spectral domain [4, 39], the captured HSI contains more complex noise than heteroscedastic Gaussian noise, as shown in Figure 2.

In this paper, we mainly focus on how to capture realistic data for learning HSI denoising network, including paired real dataset and high-quality synthetic data performing as well as real data, as shown in Figure 1. To the best of our knowledge, there is no public dataset for training and testing HSI denoising methods with diverse real world data and ground truth. Therefore, we first collect a real dataset of noisy HSI captured with short exposure time and each noisy HSI has a corresponding long-exposure clean HSI, which is beneficial to the follow-up noise model formulation and denoising method evaluation. Then, we propose an accurate noise model for HSI, which can formulate the distribution of real data well. In addition, we calibrate the parameter of the formulated noise model, and the calibrated noise model can be utilized to synthesize realistic HSI denoising dataset. Finally, we employ the generated synthetic dataset to learn a CNN for HSI denoising. Extensive experiments show that the HSI denoising network learned with only synthetic data generated with our noise model can reach the HSI denoising performance as well as that trained with real data.

In summary, our main contributions are that we

- Collect the first real dataset with paired noisy and clean HSIs, which will be publicly released to facilitate further researches;
- Formulate a noise model to synthesize realistic noisy HSI, which can match the distribution of real noisy data;
- Present a noise parameter estimation approach to calibrate the noise parameter of the given hyperspectral camera.

2. Related Work

Noise removal from a single HSI is a well-developed topic in computer vision [21] and remote sensing [30]. Ex-

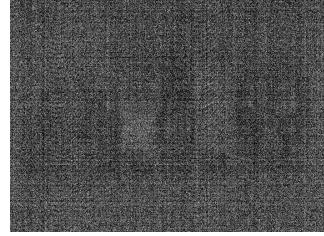


Figure 2. A typical band with the obvious noise in the HSI.

isting methods toward HSI denoising can be roughly classified into two categories, including model-based methods and learning-based methods. The model-based methods often iteratively solve an optimization problem with various hand-crafted priors, e.g., smoothness [50], sparsity [15], self-similarity [20], low-rankness [21]. Yuan *et al.* [50] employed spatial-spectral total variation to exploit smoothness prior for HSI denoising. Chen *et al.* [15] proposed HSI denoising method with sparsity prior using wavelet shrinkage and principal component analysis. Maggioni *et al.* [33] extended the BM3D filter [17] to volumetric data, which is named BM4D and explores the sparsity and self-similarity priors. Peng *et al.* [37] and Fu *et al.* [20] proposed dictionary learning methods for HSI denoising with sparsity and self-similarity priors. Xie *et al.* [46] proposed a HSI denoising method with tensor sparsity regularization. Zhang *et al.* [52] restored noisy HSI with low-rank matrix recovery. To exploit low-rankness prior for HSI denoising, more elaborately designed methods have been successively proposed [10, 11, 18, 21, 22, 43, 47]. However, the iterative optimization is time-consuming and the hand-crafted priors only model the linear property of HSI, thus, cannot sufficiently exploit the nonlinearity of various HSI in the real world.

Recently, researchers pay more attention to learning-based methods [9, 19, 29, 45, 51], which infer with less time than model-based methods leveraging graphics processing unit (GPU) and automatically learn the deep prior from training dataset. Chang *et al.* [9] proposed a HSI denoising network with residual learning and 2D convolution. Lin *et al.* [29] combined matrix factorization with deep prior for HSI denoising. Yuan *et al.* [51] employed a residual network to recover HSI with a sliding window strategy. Dong *et al.* [19] proposed a 3D HSI denoising network with Unet architecture [40] to exploit spectral and spatial correlations. Wei *et al.* [45] introduced recurrent architecture to 3D HSI denoising network to exploit global spectral correlation. Nevertheless, due to lacking paired real HSI denoising dataset, these powerful learning-based methods are often trained with synthetic data. The most widely-used approach is applying additive, white, Gaussian noise to generate noisy HSI [9, 19, 29, 51]. However, even the complex noise [16, 45] also cannot effectively model the noise in re-

alistic testing HSI, and leads to significant performance reduction according to our experiments.

Distribution difference between synthetic data and real image is a general problem in learning-based denoising methods. To alleviate this issue, there are mainly two solutions. On the one hand, some works [1, 12, 13, 26] collect paired real RGB images not only for evaluation but also for network learning. For instance, Chen *et al.* [13] collected a paired real dataset, with short-exposure noisy image and the corresponding long-exposure clean image, for RGB image denoising. Nevertheless, capturing an abundant paired real dataset is obviously expensive and requires a great amount of labor. More importantly, to the best our knowledge, there is no appropriate paired real dataset for training HSI denoising network.

On the other hand, to evade the difficulties in capturing paired real data with camera, some researches focus on improving the realism of synthetic dataset. By considering both photon and thermal noise, the works of [2] utilized a signal-dependent heteroscedastic Gaussian model to characterize the noise properties in real HSI. More recently, Chen *et al.* [16, 45] utilized a noise model, considering the Gaussian noise, stripe noise, deadline noise and impluse noise, to simulate noisy HSI. However, these methods either oversimplify noise ingredients caused by sensor, or do not estimate the noise parameter for real noisy HSI.

In this work, we collect the first dataset with short-exposure noisy HSI and corresponding long-exposure clean HSI to support systematic reproducible research in HSI denoising. Based on the captured real HSI dataset, we propose an accurate noise formulation model and a corresponding noise parameter estimation method to synthesize realistic dataset, and verify its effectiveness by comparing with the captured real HSI dataset.

3. Real HSI Denoising Dataset

The existing learning-based methods [9, 19, 29, 45, 51] often trained and evaluated on synthetic data, and the denoising and generalization capabilities on real data are not considered, as there is no appropriate paired real HSI denoising dataset.

To support the systematic research, we collect the first real dataset for training and benchmarking HSI denoising. We employ a SOC710-VP hyperspectral camera, manufactured by Surface Optics Corporation (SOC), USA. The SOC710-VP hyperspectral camera is equipped with a silicon-based charge-coupled device (CCD) and an integrated scanning system. With standard settings, the SOC710-VP can capture HSI with 696×520 pixels in spatial resolution and 256 spectral bands from 376.76 nm to 1075.80 nm at 2.7 nm interval. The camera is mounted on a sturdy tripod. For each scene, we adjust camera settings such as aperture, focus and exposure time to maximize the

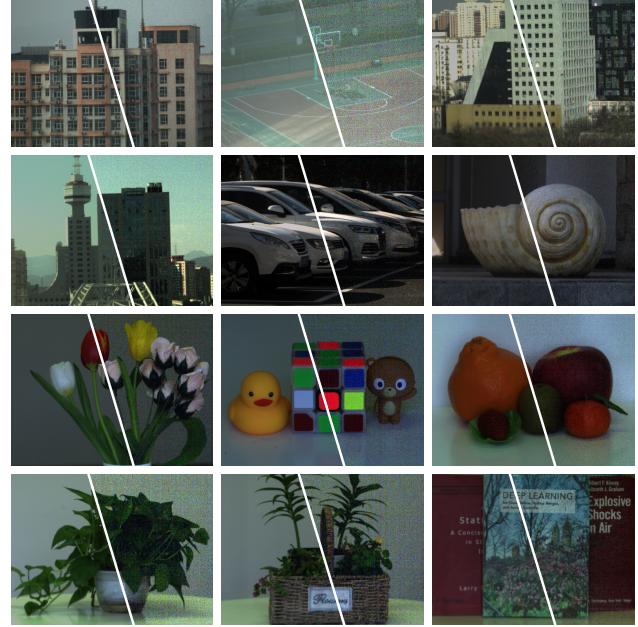


Figure 3. Example images in our captured HSI denoising dataset. Outdoor images in the top two rows, indoor images in the bottom two rows. Long-exposure clean images (ground truth) are shown in left, and short-exposure images are shown in right in each image. The noisy images are captured with $1/50$ exposure time as the reference image photographed with. The RGB is synthesized by HSI bands in 482 nm , 539 nm and 607 nm .

quality of the reference images. Capturing a reference image often needs dozens of seconds. Then, we employ a remote control software to deliberately decrease the exposure time by factor 50 for short-exposure image. In other word, we capture noisy image with fractions of a second to seconds. Since we capture multiple images for one scene and the exposure time for reference image is necessarily long, all scenes in the dataset are static. Some samples of reference images and corresponding noisy images are shown in Figure 3. The dataset contains both indoor and outdoor scenes. To capture high-quality reference images, we capture indoor images under incandescent lamp to lighten the scene. The outdoor images are generally captured on sunny days between 9:00 and 17:00 around, when the scene is well-illuminated.

Following existing HSI datasets [3, 8, 49], we select 34 bands around from 400 nm to 700 nm in visible spectral range. The dynamic range of the captured HSI is 12 bit, and the spectral value range is 0 to 4095. This dataset contains 62 long-exposure clean HSIs, of which each is paired with a corresponding noisy HSI captured with short-exposure time. We select paired noisy and reference images of 17 scenes to form the testing set, and the rest are selected for training set. The captured real HSI denoising dataset can evaluate the generalization capability of exist-

ing learning-based methods and verify the effectiveness of our noise model proposed in the next section.

4. Noise Modeling for HSI

Existing learning-based HSI denoising methods [9, 19, 29, 45, 51] are always learned on synthetic dataset, which follows various noise models. However, there is no systematic and quantitative evaluation on real testing data for these noise models.

In this section, we focus on synthesizing a realistic HSI denoising dataset. We first formulate the noise model of HSI. Then, we introduce a noise parameter estimation approach to calibrate the noise model. Finally, we synthesize a realistic dataset for learning HSI denoising network based on calibrated noise model.

4.1. Noise Formulation

Nowadays, most hyperspectral cameras utilize CCD sensors. Fortunately, some works have discussed the various noise components associated with CCD camera systems [23, 24, 41]. Let $\mathcal{X}(m, n, \lambda)$ indicate the clean HSI, where $1 \leq m \leq M$ and $1 \leq n \leq N$ index the spatial coordinate and $1 \leq \lambda \leq \Lambda$ indexes the spectral coordinate. A linear model describing the relationship between the raw sensor output in digital numbers, *i.e.*, noisy HSI, and the integrated photoelectrons during exposure can be expressed as

$$\begin{aligned} \mathcal{Y}(m, n, \lambda) &= \mathcal{X}(m, n, \lambda) + \mathcal{N}(m, n, \lambda) \\ &= k\mathcal{L}(m, n, \lambda) + \mathcal{N}(m, n, \lambda), \end{aligned} \quad (1)$$

where \mathcal{Y} is the noisy HSI, \mathcal{N} indicates the summation of all noises physically caused by light and camera, \mathcal{L} represents the number of photoelectrons that is proportional to the scene irradiation, and k denotes the system gain, respectively. Note that we assume the system gain of all elements in the captured HSI is identical due to the same CCD sensor in hyperspectral camera.

To systematically analyze the noise in HSI, we divide the noise into two components, *e.g.*, signal-dependent noise (correlated to incident light) and signal-independent noise (uncorrelated to incident light).

Signal-dependent noise. During the exposure time, photons in the incident light hits the sensing area of the sensor. Leveraging photoelectric conversion, the sensor transforms photon to electrons. However, due to the quantum property of light, the number of electrons collected by sensor exists an ineluctable uncertainty, which can be formulated as a Poisson distribution

$$[\mathcal{L}(m, n, \lambda) + \mathcal{N}_{sd}(m, n, \lambda)] \sim p(\mathcal{L}(m, n, \lambda)), \quad (2)$$

where \mathcal{N}_{sd} denotes the signal-dependent noise (*i.e.*, shot noise) and $p(\cdot)$ represents the Poisson distribution. Generally, due to the limitation of light characteristic, shot noise

cannot be avoided by hyperspectral sensors. Overall, signal-dependent noise is shot noise caused by photon-to-electrons stage.

Signal-independent noise. For CCD sensors in hyperspectral imaging, during the exposure time, thermal energy in silicon generates free electrons, known as dark current noise, which can be stored at collection sites and thereafter become indistinguishable from photoelectrons. Read noise is caused by amplifier, reset, and other electronic noise sources during electrons-to-voltage stage. During voltage-to-digital stage, due to the dynamic range of digital storage medium, the continuous analog voltage signal is quantized to discrete digital number, which causes the quantization noise. These noises can be described as

$$\mathcal{N}_p(m, n, \lambda) = \mathcal{N}_d(m, n, \lambda) + \mathcal{N}_r(m, n, \lambda) + \mathcal{N}_q(m, n, \lambda), \quad (3)$$

where \mathcal{N}_d is the dark current noise, \mathcal{N}_r is read noise, and \mathcal{N}_q is quantization noise. As these noises are caused by camera circuit and not related to incident light, they are signal-independent noises. Following the zero-mean noise assumption for all pixels, these signal-independent noise can be formulated as a Gaussian distribution

$$\mathcal{N}_p(n, m, \lambda) \sim g(0, \sigma_p(\lambda)), \quad (4)$$

where $g(\cdot)$ represents the Gaussian distribution and $\sigma_p(\lambda)$ denotes the scale parameter for the λ -th band.

Apart from these signal-independent noise, due to spatial scanning design, hyperspectral camera often suffers from stripe pattern noise, as illustrated in Figure 2. As we analyze the hyperspectral camera employed to capture our real HSI dataset, we find it contains both horizontal stripe pattern noise and vertical stripe pattern noise, as shown in Figure 4. The reason is that each row is captured by the same CCD unit during horizontal scanning, which causes the horizontal stripe pattern noise. Meanwhile, each column is captured in different time, which causes the vertical stripe pattern noise. Thus, the **stripe pattern noise** \mathcal{N}_{sp} can be represented as

$$\mathcal{N}_{sp}(n, m, \lambda) = \mathcal{N}_{sp}^h(n, m, \lambda) + \mathcal{N}_{sp}^v(n, m, \lambda), \quad (5)$$

where \mathcal{N}_{sp}^h and \mathcal{N}_{sp}^v denote horizontal and vertical stripe pattern noise, respectively. The stripe pattern noise is caused by scanning camera design and not related to incident light, and it is signal-independent noise. Following the assumption of zero-mean Gaussian distribution in each row or column, we respectively formulate the horizontal and vertical stripe pattern noise as

$$\begin{aligned} \mathcal{N}_{sp}^h(m, \lambda) &\sim g(0, \sigma_h(\lambda)), \\ \mathcal{N}_{sp}^v(n, \lambda) &\sim g(0, \sigma_v(\lambda)), \end{aligned} \quad (6)$$

where $\sigma_h(\lambda)$ and $\sigma_v(\lambda)$ denote the scale parameters of horizontal and vertical stripe pattern noise for the λ -th band,

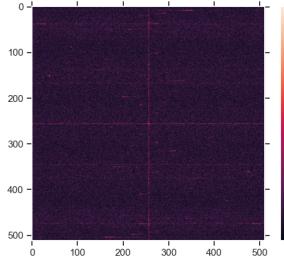


Figure 4. Analysis of stripe pattern noise. Performing discrete Fourier transform on a bias frame, the highlighted vertical and horizontal patterns in the centralized Fourier spectrum verify the existence of both horizontal and vertical stripe pattern noises, respectively.

respectively. Thus, the full signal-independent noise can be described as

$$\mathcal{N}_{si} = \mathcal{N}_p + \mathcal{N}_{sp}. \quad (7)$$

To summarize, our full noise model can be formulated as

$$\begin{aligned} \mathcal{N} &= k\mathcal{N}_{sd} + \mathcal{N}_{si}, \\ &= k\mathcal{N}_{sd} + \mathcal{N}_p + \mathcal{N}_{sp}, \end{aligned} \quad (8)$$

where k , \mathcal{N}_{sd} and \mathcal{N}_{si} represents system gain, signal-dependent noise and signal-independent noise, respectively.

4.2. Noise Parameter Estimation

To calibrate the noise parameters of our proposed noise model for the given hyperspectral camera, we introduce a noise parameter estimation method. According to Equations (2) (4) and (6), four parameters need to be calibrated, including the system gain k for signal-dependent noise \mathcal{N}_{sd} , the scale parameters σ_p for pixel noise \mathcal{N}_p , and scale parameters σ_h and σ_v for stripe pattern noise \mathcal{N}_{sp} .

To estimate the system gain k , we record a colorchecker under uniform light, as shown in Figure 5. Each color block in colorchecker is cropped to form a sequence of images captured with different intensities. Thus, only one image need to be captured rather than multiple images with different luminous fluxes. According to the Photon Transfer approach [25], we employ the sequence of color blocks captured under well-lighted environment to determine system gain k . With the estimated system gain k , we can add signal-dependent noise to HSI.

To estimate the scale parameters σ_p , σ_h and σ_v , the bias image is captured with the shortest exposure time under a lightless condition, *i.e.*, in a dark room and capping on the camera lens. Bias image delineates the noise picture independent of light, blended by the multiple noise sources aforementioned. Leveraging the zero-mean assumption of pixel noise \mathcal{N}_p , we firstly extract the mean values of each column or row of the λ -th band in bias image to estimate

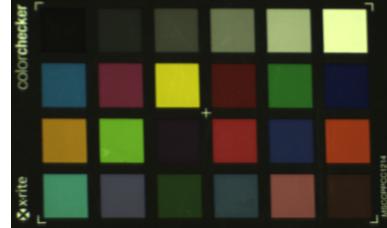


Figure 5. The colorchecker employed to estimate system gain k .

the underlying intensities of vertical or horizontal stripe pattern noise. Then, we can readily estimate the scale parameters $\sigma_h(\lambda)$ or $\sigma_v(\lambda)$ by approximating Gaussian distribution. Further, we subtract the estimated stripe pattern noise from bias image and estimate the scale parameters $\sigma_p(\lambda)$ of pixel noise by maximizing the log-likelihood of Gaussian distribution. Figure 6 shows the probability plot. The goodness-of-fit can be evaluated by the coefficient of determination R^2 [34], and the R^2 is closer to 1 illustrates the fitness is better. We can see that the R^2 of $\sigma_p(\lambda)$, $\sigma_h(\lambda)$ and $\sigma_v(\lambda)$ are all close to 1, which means our method fits the empirical data well.

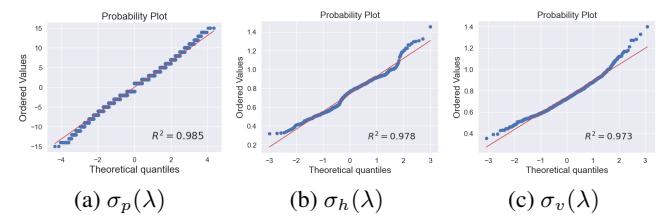


Figure 6. The distribution fitting of signal-independent noise (a), horizontal (b) and vertical (c) stripe noises. $\lambda = 550$ nm.

To make our calibrated noise model more robust, we estimate scale parameters of a sequence of bias images, and fit them with Gaussian distribution in the logarithmic domain. When we employ our calibrated noise model, we can sample the noise parameter from

$$\begin{aligned} \log(\sigma_p(\lambda)) &\sim g(a_p(\lambda), b_p(\lambda)), \\ \log(\sigma_h(\lambda)) &\sim g(a_h(\lambda), b_h(\lambda)), \\ \log(\sigma_v(\lambda)) &\sim g(a_v(\lambda), b_v(\lambda)), \end{aligned} \quad (9)$$

where $a(\lambda)$ and $b(\lambda)$ denote the estimated mean and standard deviation of Gaussian distribution for λ -th band, respectively.

4.3. Noisy Image Synthesis

Here, we describe how to synthesize noisy image with the calibrated noise model. We first divide the clean HSI by a factor to match the intensity of short-exposure noisy image, which follows the setting of our captured real HSI denoising dataset in Section 3. Then, we add signal-dependent noise \mathcal{N}_{sd} by converting the HSI \mathcal{X} into the number of photoelectrons \mathcal{L} , imposing a Poisson distribution on \mathcal{L} , and

Table 1. Quantitative results of different methods on our real HSI denoising dataset. Ours(Real) and Ours(Synthetic) denote our network trained with our real and synthetic datasets, respectively. The best results are highlighted in **bold**.

Metrics	Methods						
	Noisy	BM4D [33]	ITSReg [46]	LRTDTV [43]	QRNN3D [45]	Ours (Real)	Ours (Synthetic)
PSNR	20.907	25.318	25.459	25.564	23.832	30.583	30.451
SSIM	0.3186	0.8156	0.8400	0.7859	0.7917	0.8843	0.9004
SAM	25.2992	6.3024	5.1425	6.4881	10.0188	4.6816	4.1364
ERGAS	60.758	35.078	34.618	34.429	43.271	20.816	21.324

reverting the signal to \mathcal{X} , with the estimated system gain k . Besides, we sample the noise parameters according to Equation (9), and generate pixel noise \mathcal{N}_p and stripe pattern noise \mathcal{N}_{sp} . These generated noise are added to the scaled clean HSI. Finally, we multiply the noisy HSI by the same factor to match the intensity of clean HSI.

Following these steps, we generate a realistic synthetic dataset with rich paired noisy and clean HSIs, which is beneficial to train HSI denoising network and generalize it to the evaluation on noisy HSI in the real world.

5. Experimental Results

In this section, we first introduce the settings in our experiments, including implementation details and metrics for quantitative evaluation. Then, our method is compared with several state-of-the-art methods on our captured real HSI denoising dataset. Finally, we discuss the effect of different noise model to synthesize noisy data.

5.1. Settings

Implementation details. Recent researches [13, 31, 48] have shown the effectiveness of CNN in image-to-image tasks, especially U-Net [40]. Inspired by it, we employ a modified U-Net with 3D convolution to better exploit spectral correlation¹.

To train our HSI denoising network, we crop overlapped 256×256 spatial regions from realistic dataset synthesized with our noise model and augment them by random flipping and/or rotation. Besides, we also train networks with our real HSI denoising dataset and employ the same augmentation method.

Our implementation is based on PyTorch [36]. The models are trained with the L_1 loss and Adam optimizer [27] ($\beta_1 = 0.9$ and $\beta_2 = 0.999$) for 200 epochs. The initial learning rate and mini-batch size are set to 2×10^{-4} and 4, respectively.

Evaluation metrics. We employ four quantitative quality metrics to evaluate the performance of all methods, including peak signal-to-noise ratio (PSNR), structural similarity

¹The network design is not the contribution of this paper, and can be replaced by other powerful HSI denoising network.

(SSIM) [44], spectral angle mapping (SAM) [28] and relative dimensionless global error in synthesis (ERGAS) [42]. PSNR and SSIM show the spatial accuracy, which are calculate on each 2D spatial image and averaged over all spectral band. SAM and ERGAS show the spectral fidelity. SAM is calculated on each 1D spectral vector and averaged over all spatial points, while ERGAS calculates the average amount of specific spectral distortion normalized by mean intensity of each band. Larger values of PSNR and SSIM indicate better reconstruction, and smaller values of SAM and ERGAS denote higher performance.

5.2. Evaluation on Real HSI Denoising Dataset

Compared methods. We compare our method with four state-of-the-art HSI denoising methods on our captured real HSI dataset, including three model-based methods, *i.e.*, filtering-based approach (BM4D [33]), sparsity-based approach (ITSReg [46]), low-rankness tensor method (LRTDTV [43]), and one learning-based method, *i.e.*, QRNN3D [45]. We make great effort to reproduce the best results for competitive methods with the codes that are released publicly. Note that to investigate the generalization capability of existing learning-based methods, we employ QRNN3D trained with synthetic complex noise² to train and evaluate on our real HSI denoising dataset.

Numerical results. Table 1 provides the averaged recovery results over all test images on our real HSI denoising dataset with indoor and outdoor scenes, to quantitatively compare our method with BM4D, ITSReg, LRTDTV and QRNN3D. The best results are highlighted in bold for each metric. It can be seen that our method always has better performance than all compared methods, which demonstrates the effectiveness of our method. Compared with the results of QRNN3D, our model trained with real and synthetic datasets both significantly outperforms QRNN3D learned on our dataset with its complex noise model, even if QRNN3D is elaborately designed and has more complex computation. It implies the importance of realistic training dataset to guarantee the generalization capability for image in the real world. *Moreover, our network trained with synthetic dataset is comparable with or even better than that trained with real dataset in some metrics.* The reason may

²The noise model in [45] is the same with that in [16].

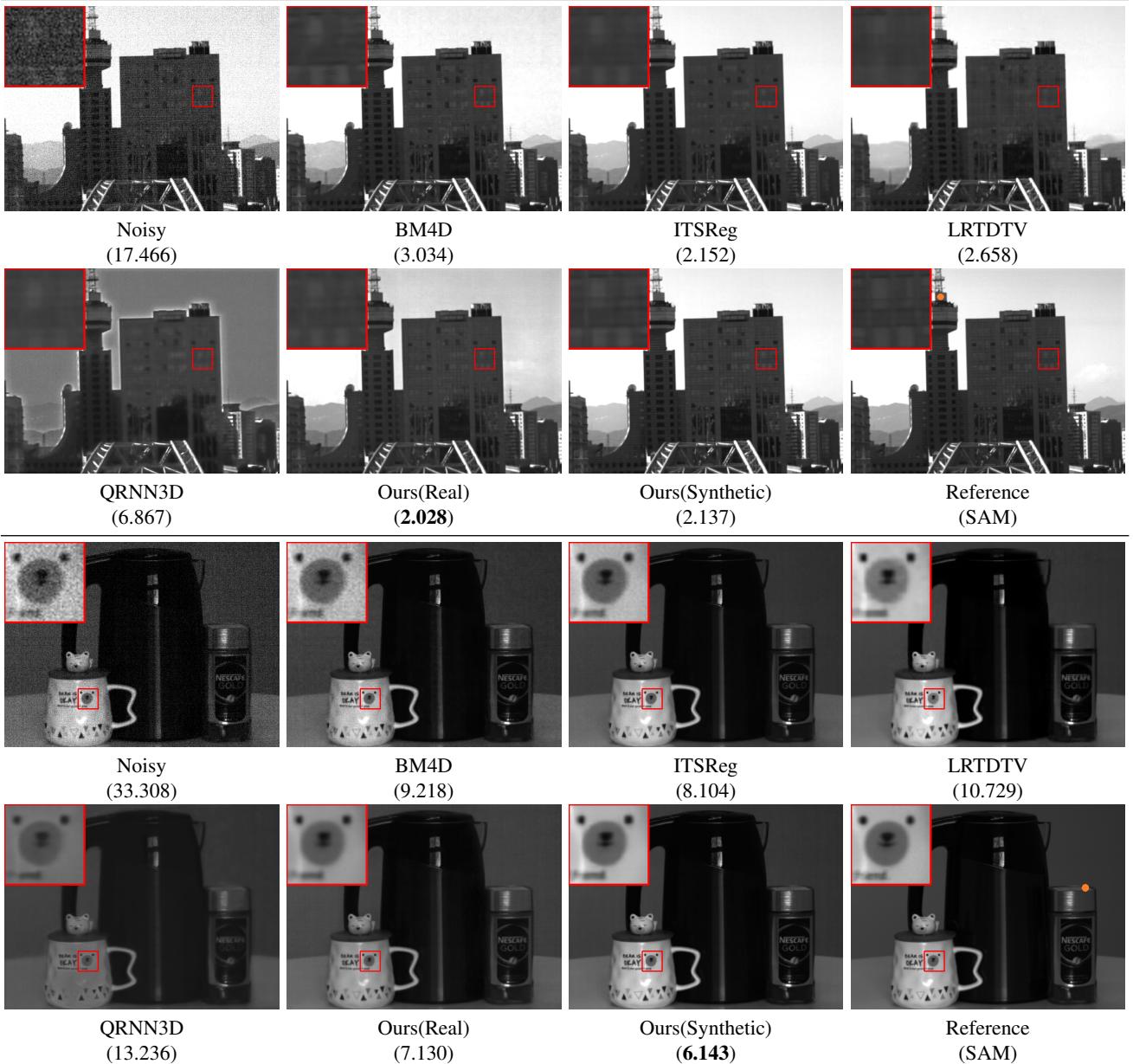


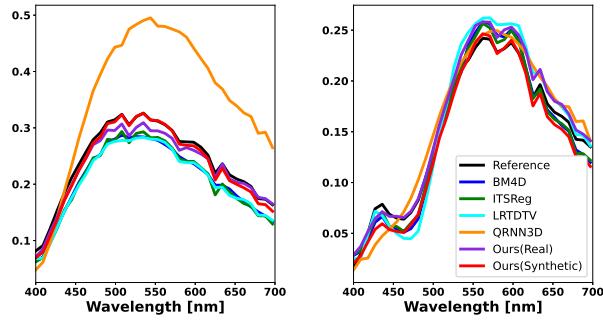
Figure 7. Visual quality comparison on two typical scenes for noise removing in our real HSI denoising dataset. The outdoor and indoor scenes are shown in top and bottom rows, respectively. The noisy image, BM4D/ITSReg/LRTDTV/QRNN3D/Ours(Real)/Ours(Synthetic) recovered results and the reference image on 550 nm are shown from left to right and from top to bottom. The SAM for the result images is shown in the parenthesis. Our methods outperform all the competitive methods in terms of spatial and spectral fidelities.

be that the synthetic dataset generated with our noise model can provide more samples for the same number of scenes, compared with real dataset. It further verifies the superiority of our accurately noise model.

Spatial quality. To visualize the experimental results, two representative scenes are shown in Figure 7. The PSNR and SAM values are provided for each results. Compared with other methods, the recovered results by our method are consistently more accurate for all scenes. Specifically, our

method can produce the visually pleasant results with less artifacts and sharp edges. It verifies that our method can provide higher spatial accuracy.

Spectral fidelity. Figure 8 shows the recovered spectra of two points in the selected images, which is indicated as orange in Figure 7. We can see that the spectra recovered by our method are much closer to the reference, which demonstrates that our method obtain higher spectral fidelity.



Outdoor

Indoor

Figure 8. Comparison of spectral fidelity. The points of outdoor and indoor scenes are indicated as orange in Figure 7. The spectra recovered by our method are much closer to the reference.

Table 2. Quantitative results of our network trained with synthetic datasets generated by different noise models. The complex noise model consists of Gaussian, stripe, deadline and impulse noises, and is utilized in previous learning-based method [45]. Homo. denotes the homoscedastic Gaussian noise model for pixel noise (\mathcal{N}_p) and Hetero. represents the heteroscedastic Gaussian noise model for signal-dependent and pixel noises ($\mathcal{N}_{sd} + \mathcal{N}_p$). The best results are highlighted in **bold**.

Noise models	Metrics			
	PSNR	SSIM	SAM	ERGAS
Complex [16]	28.337	0.8327	7.9049	28.831
Homo.	29.976	0.8446	5.8650	22.505
Hetero.	30.284	0.8946	4.4488	21.675
Ours	30.451	0.9004	4.1364	21.324

5.3. Ablation Study on Noise Models

To verify the effectiveness of the proposed noise model for HSI, we compare the performance of network trained with synthetic dataset generated by different noise models, including complex noise model [16], homoscedastic Gaussian noise model and heteroscedastic Gaussian noise model [2]. The results are provided in Table 2. The complex noise model is utilized by previous learning-based method [45] to synthesize dataset. However, the complex noise model performs worse than latter three calibrated noise models, for it does not well calibrate and match the noise distribution of real HSI data. Homoscedastic and heteroscedastic Gaussian noise models perform worse than our full noise model, even if we have calibrated the noise parameters for them. A visual comparison of our noise model and other noise models is provided in Figure 9. It can be seen that the result of our method is cleaner than compared methods, which verifies that our method can accurately formulate the noise in the HSI and can generate the most realistic dataset for CNN training.

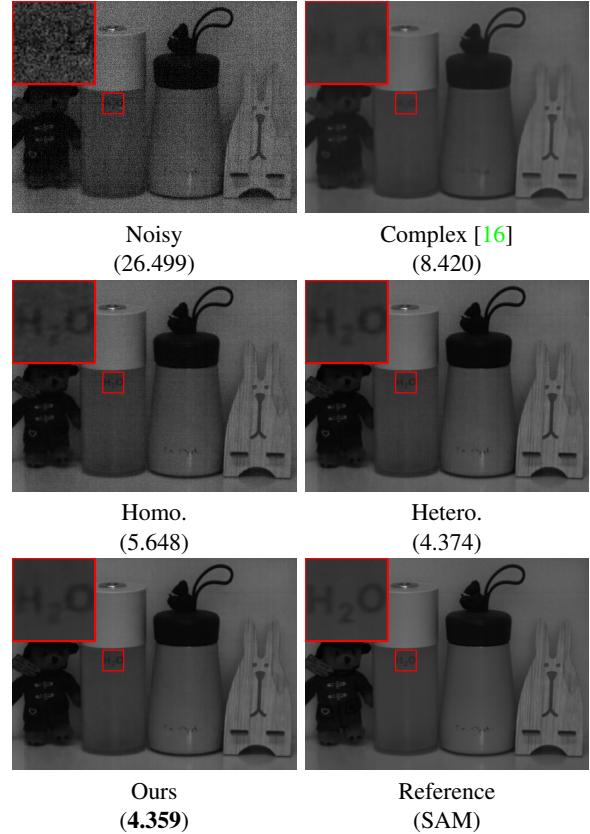


Figure 9. Visual quality comparison of our network trained with different noise models. We show the spectral band in 550 nm.

6. Conclusion

In this paper, we mainly focus on how to capture realistic datasets for learning HSI denoising network. On the one hand, we collect the first real HSI denoising dataset with paired short-exposure noisy image and long-exposure clean image, which is beneficial to follow-up researches in this area. On the other hand, we propose an accurate noise model to comprehensively consider the noise occurred in imaging process. Besides, we present a noise parameter estimation approach to calibrate the proposed noise model for a given hyperspectral camera, and the calibrated noise model can be employed to synthesize realistic dataset. Experimental results show that a network learned with only synthetic data generated with our noise model performs as well as that learned with real data. We hope our work could provide foundations for further researches in the field of HSI denoising with realistic data.

Acknowledgments

This work was supported by the National Natural Science Foundation of China under Grants No. 62171038, No. 61827901 and No. 62088101.

References

- [1] Abdelrahman Abdelhamed, Stephen Lin, and Michael S. Brown. A high-quality denoising dataset for smartphone cameras. In *Proc. of Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018. 2, 3
- [2] Nicola Acito, Marco Diani, and Giovanni Corsini. Signal-dependent noise modeling and model parameter estimation in hyperspectral images. *IEEE Trans. Geoscience and Remote Sensing*, 49(8):2957–2971, 2011. 2, 3, 8
- [3] Boaz Arad and Ohad Ben-Shahar. Sparse recovery of hyperspectral signal from natural rgb images. In *Proc. of European Conference on Computer Vision (ECCV)*, 2016. 3
- [4] Robert W. Basedow, Dwayne C. Carmer, and Mark E. Anderson. Hydice system: Implementation and performance. In *Proc. of SPIE's Symposium on OE/Aerospace Sensing and Dual Use Photonics*, pages 258–267, 1995. 1, 2
- [5] Asgeir Bjorgan and Lise Lyngsnes Randeberg. Towards real-time medical diagnostics using hyperspectral imaging technology. In *Proc. of Clinical and Biomedical Spectroscopy and Imaging IV*, page 953712, 2015. 1
- [6] Marcus Borengasser, William S. Hungate, and Russell Watkins. *Hyperspectral Remote Sensing: Principles and Applications*. Remote Sensing Applications Series. CRC Press, Dec. 2007. 1
- [7] Xiangyong Cao, Feng Zhou, Lin Xu, Deyu Meng, Zongben Xu, and John Paisley. Hyperspectral image classification with markov random fields and a convolutional neural network. *IEEE Trans. Image Processing*, 27(5):2354–2367, 2018. 1
- [8] Ayan Chakrabarti and Todd E. Zickler. Statistics of real-world hyperspectral images. In *Proc. of Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 193–200, 2011. 3
- [9] Yi Chang, Luxin Yan, Houzhang Fang, Sheng Zhong, and Wenshan Liao. Hsi-denet: Hyperspectral image restoration via convolutional neural network. *IEEE Trans. Geoscience and Remote Sensing*, 57(2):667–682, 2018. 1, 2, 3, 4
- [10] Yi Chang, Luxin Yan, Xi-Le Zhao, Houzhang Fang, Zhijun Zhang, and Sheng Zhong. Weighted low-rank tensor recovery for hyperspectral image restoration. *IEEE Trans. Cybernetics*, 50(11):4558–4572, 2020. 2
- [11] Yi Chang, Luxin Yan, and Sheng Zhong. Hyper-laplacian regularized unidirectional low-rank tensor recovery for multispectral image denoising. In *Proc. of Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4260–4268, 2017. 2
- [12] Chen Chen, Qifeng Chen, Minh N Do, and Vladlen Koltun. Seeing motion in the dark. In *Proc. of International Conference on Computer Vision (ICCV)*, pages 3185–3194, 2019. 2, 3
- [13] Chen Chen, Qifeng Chen, Jia Xu, and Vladlen Koltun. Learning to see in the dark. In *Proc. of Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3291–3300, 2018. 2, 3, 6
- [14] Chen Chen, Wei Li, Eric W Tramel, Minshan Cui, Saurabh Prasad, and James E Fowler. Spectral–spatial preprocessing using multihypothesis prediction for noise-robust hyperspectral image classification. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 7(4):1047–1059, 2014. 1
- [15] Guangyi Chen and Shen-En Qian. Denoising of hyperspectral imagery using principal component analysis and wavelet shrinkage. *IEEE Trans. Geoscience and Remote Sensing*, 49(3):973–980, 2010. 2
- [16] Yang Chen, Xiangyong Cao, Qian Zhao, Deyu Meng, and Zongben Xu. Denoising hyperspectral image with non-iid noise structure. *IEEE Trans. Cybernetics*, 48(3):1054–1066, 2017. 1, 2, 3, 6, 8
- [17] Kostadin Dabov, Alessandro Foi, Vladimir Katkovnik, and Karen Egiazarian. Image denoising by sparse 3-d transform-domain collaborative filtering. *IEEE Trans. Image Processing*, 16(8):2080–2095, 2007. 2
- [18] Weisheng Dong, Guangyu Li, Guangming Shi, Xin Li, and Yi Ma. Low-rank tensor approximation with laplacian scale mixture modeling for multiframe image denoising. In *Proc. of International Conference on Computer Vision (ICCV)*, pages 442–449, 2015. 2
- [19] Weisheng Dong, Huan Wang, Fangfang Wu, Guangming Shi, and Xin Li. Deep spatial–spectral representation learning for hyperspectral image denoising. *IEEE Trans. Computational Imaging*, 5(4):635–648, 2019. 1, 2, 3, 4
- [20] Ying Fu, Antony Lam, Imari Sato, and Yoichi Sato. Adaptive spatial-spectral dictionary learning for hyperspectral image restoration. *International Journal of Computer Vision*, 122(2):228–245, 2017. 1, 2
- [21] Wei He, Quanning Yao, Chao Li, Naoto Yokoya, and Qibin Zhao. Non-local meets global: An integrated paradigm for hyperspectral denoising. In *Proc. of Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 6868–6877, 2019. 1, 2
- [22] Wei He, Hongyan Zhang, Liangpei Zhang, and Huanfeng Shen. Total-variation-regularized low-rank matrix factorization for hyperspectral image restoration. *IEEE Trans. Geoscience and Remote Sensing*, 54(1):178–188, 2015. 2
- [23] Glenn E Healey and Raghava Kondepudy. Radiometric ccd camera calibration and noise estimation. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 16(3):267–276, 1994. 4
- [24] Gerald C Holst. Ccd arrays, cameras, and displays. 1998. 4
- [25] James Janesick, Kenneth Klaasen, and Tom Elliott. Ccd charge collection efficiency and the photon transfer technique. In *Proc. of Solid-state imaging arrays*, volume 570, pages 7–19, 1985. 5
- [26] Haiyang Jiang and Yinqiang Zheng. Learning to see moving objects in the dark. In *Proc. of International Conference on Computer Vision (ICCV)*, pages 7324–7333, 2019. 2, 3
- [27] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014. 6
- [28] Fred A Kruse, AB Lefkoff, JW Boardman, KB Heidebrecht, AT Shapiro, PJ Barloon, and AFH Goetz. The spectral image processing system (sips) interactive visualization and analysis of imaging spectrometer data. *Remote sensing of environment*, 44(2-3):145–163, 1993. 6

- [29] Baihong Lin, Xiaoming Tao, and Jianhua Lu. Hyperspectral image denoising via matrix factorization and deep prior regularization. *IEEE Trans. Image Processing*, 29:565–578, 2019. [1](#), [2](#), [3](#), [4](#)
- [30] Xuefeng Liu, Salah Bourennane, and Caroline Fossati. Denoising of hyperspectral images using the parafac model and statistical performance analysis. *IEEE Trans. Geoscience and Remote Sensing*, 50(10):3717–3724, 2012. [2](#)
- [31] Jonathan Long, Evan Shelhamer, and Trevor Darrell. Fully convolutional networks for semantic segmentation. In *Proc. of Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3431–3440, 2015. [6](#)
- [32] Guolan Lu and Baowei Fei. Medical hyperspectral imaging: a review. *Journal of Biomedical Optics*, 19(1):010901, 2014. [1](#)
- [33] Matteo Maggioni, Vladimir Katkovnik, Karen Egiazarian, and Alessandro Foi. Nonlocal transform-domain filter for volumetric data denoising and reconstruction. *IEEE Trans. Image Processing*, 22(1):119–133, 2012. [2](#), [6](#)
- [34] Eugene C Morgan, Matthew Lackner, Richard M Vogel, and Laurie G Baise. Probability distributions for offshore wind speeds. *Energy Conversion and Management*, 52(1):15–26, 2011. [5](#)
- [35] Lujendra Ojha, Mary Beth Wilhelm, Scott L. Murchie, Alfred S. McEwen, James J. Wray, Jennifer Hanley, Marion Mass, and Matt Chojnacki. Spectral evidence for hydrated salts in recurring slope lineae on Mars. *Nature Geoscience*, 8(11):829–832, Nov. 2015. [1](#)
- [36] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, et al. Pytorch: An imperative style, high-performance deep learning library. *arXiv preprint arXiv:1912.01703*, 2019. [6](#)
- [37] Yi Peng, Deyu Meng, Zongben Xu, Chenqiang Gao, Yi Yang, and Biao Zhang. Decomposable nonlocal tensor dictionary learning for multispectral image denoising. In *Proc. of Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2949–2956, 2014. [2](#)
- [38] Tobias Plotz and Stefan Roth. Benchmarking denoising algorithms with real photographs. In *Proc. of Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1586–1595, 2017. [2](#)
- [39] Wallace M. Porter and Harry T. Enmark. A system overview of the airborne visible/infrared imaging spectrometer (aviris). In *Proc. of Annual Technical Symposium*, pages 22–31, 1987. [1](#), [2](#)
- [40] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *Proc. of International Conference on Medical image computing and computer-assisted intervention*, pages 234–241, 2015. [2](#), [6](#)
- [41] John R Schott. *Remote sensing: the image chain approach*. Oxford University Press on Demand, 2007. [4](#)
- [42] Lucien Wald. Quality of high resolution synthesised images: Is there a simple criterion? In *Proc. of Conference on Fusion of Earth Data*, pages 99–103, 2000. [6](#)
- [43] Yao Wang, Jiangjun Peng, Qian Zhao, Yee Leung, Xi-Le Zhao, and Deyu Meng. Hyperspectral image restoration via total variation regularized low-rank tensor decomposition. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 11(4):1227–1243, 2017. [2](#), [6](#)
- [44] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE Trans. Image Processing*, 13(4):600–612, 2004. [6](#)
- [45] Kaixuan Wei, Ying Fu, and Hua Huang. 3-d quasi-recurrent neural network for hyperspectral image denoising. *IEEE Trans. Neural Networks and Learning Systems*, 32(1):363–375, 2021. [1](#), [2](#), [3](#), [4](#), [6](#), [8](#)
- [46] Qi Xie, Qian Zhao, Deyu Meng, Zongben Xu, Shuhang Gu, Wangmeng Zuo, and Lei Zhang. Multispectral images denoising by intrinsic tensor sparsity regularization. In *Proc. of Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1692–1700, 2016. [2](#), [6](#)
- [47] Yuan Xie, Yanyun Qu, Dacheng Tao, Weiwei Wu, Qiangqiang Yuan, and Wensheng Zhang. Hyperspectral image restoration via iteratively regularized weighted schatten p -norm minimization. *IEEE Trans. Geoscience and Remote Sensing*, 54(8):4642–4659, 2016. [2](#)
- [48] Li Xu, Jimmy Ren, Qiong Yan, Renjie Liao, and Jiaya Jia. Deep edge-aware filters. In *Proc. of International Conference on Machine Learning (ICML)*, pages 1669–1678, 2015. [6](#)
- [49] Fumihiro Yasuma, Tommo Mitsunaga, Daisuke Iso, and Shree K. Nayar. Generalized assorted pixel camera: Post-capture control of resolution, dynamic range and spectrum. *IEEE Trans. Image Processing*, 19(9):2241–2253, 2010. [3](#)
- [50] Qiangqiang Yuan, Liangpei Zhang, and Huanfeng Shen. Hyperspectral image denoising employing a spectral–spatial adaptive total variation model. *IEEE Trans. Geoscience and Remote Sensing*, 50(10):3660–3677, 2012. [1](#), [2](#)
- [51] Qiangqiang Yuan, Qiang Zhang, Jie Li, Huanfeng Shen, and Liangpei Zhang. Hyperspectral image denoising employing a spatial–spectral deep residual convolutional neural network. *IEEE Trans. Geoscience and Remote Sensing*, 57(2):1205–1218, 2018. [1](#), [2](#), [3](#), [4](#)
- [52] Hongyan Zhang, Wei He, Liangpei Zhang, Huanfeng Shen, and Qiangqiang Yuan. Hyperspectral image restoration using low-rank matrix recovery. *IEEE Trans. Geoscience and Remote Sensing*, 52(8):4729–4743, 2013. [2](#)