

# A Comprehensive Comparison of Multi-Dimensional Image Denoising Methods

Zhaoming Kong, Xiaowei Yang, and Lifang He

**Abstract**—Filtering multi-dimensional images such as color images, color videos, multispectral images and magnetic resonance images is challenging in terms of both effectiveness and efficiency. Leveraging the nonlocal self-similarity (NLSS) characteristic of images and sparse representation in the transform domain, the block-matching and 3D filtering (BM3D) based methods show powerful denoising performance. Recently, numerous new approaches with different regularization terms, transforms and advanced deep neural network (DNN) architectures are proposed to improve denoising quality. In this paper, we extensively compare over 60 methods on both synthetic and real-world datasets. We also introduce a new color image and video dataset for benchmarking, and our evaluations are performed from four different perspectives including quantitative metrics, visual effects, human ratings and computational cost. Comprehensive experiments demonstrate: (i) the effectiveness and efficiency of the BM3D family for various denoising tasks, (ii) a simple matrix-based algorithm could produce similar results compared with its tensor counterparts, and (iii) several DNN models trained with synthetic Gaussian noise show state-of-the-art performance on real-world color image and video datasets. Despite the progress in recent years, we discuss shortcomings and possible extensions of existing techniques. Datasets and codes for evaluation are made publicly available at <https://github.com/ZhaomingKong/Denoising-Comparison>.

**Index Terms**—Multi-dimensional image denoising, nonlocal self-similarity, transform-domain techniques, block-matching filters, deep neural network.

## 1 INTRODUCTION

Image denoising plays an important role in modern imaging systems, and it has attracted much attention from both academia and industry. Theoretically, image denoising is a special case of inverse problems [1], which aims to estimate the underlying clean image from the noisy observation. In real-world applications, it can be used as a pre-processing step for subsequent tasks, such as visual tracking [2], segmentation [3] and classification [4]. Also, it may serve as the ultimate goal of image quality enhancement to produce more visually pleasant images.

Image denoising enjoys a rich history and the earliest works may date back to the 1950s [5]. The recent surge of denoising methods is mainly credited to the famous block-matching 3D (BM3D) [6] framework, which combines the nonlocal similarity characteristic [7] of natural images and the sparse representation in the transform domain [8]. At an early stage, many related works focus on filtering single-channel grayscale images [9], [10]. Recently, the advancement of imaging systems and technologies has largely enriched the information preserved and presented by multi-dimensional images, which could deliver more faithful representation for real scenes. For example, the popularity of mobile camera phones facilitates the capturing and recording of colorful and dynamic objects. The multispectral imaging (MSI) and hyperspectral imaging (HSI) techniques provide image information in both the spatial and spectral domain. Magnetic resonance imaging (MRI) utilizes a non-invasive imaging technology that produces

three dimensional detailed anatomical images often used for disease detection, diagnosis, and treatment monitoring.

The growth of image size and dimension also posts greater demands on denoising. One major challenge of handling higher dimensional images is to effectively exploit the inter-correlated information of multiple channels or spectrums, and also find a balance between noise removal and detail preservation. In the past two decades, the representative BM3D method has been successfully extended to multi-dimensional images in two different ways. The first strategy is to apply a certain decorrelation transform such that in the transformed space each channel or spectral band may be filtered independently by some effective single-channel denoiser. For example, the color-BM3D (CBM3D) method [11] shows that the established opponent or YCbCr color transform of natural RGB images could provide near-optimal decorrelation of the color data [12]. An alternative solution is to model and utilize the channel- or band-wise correlation by processing the full set of the multi-dimensional image data jointly. To achieve this goal, Maggioni et al. [13] extend BM3D to BM4D by using 3D cubes of voxels, which are then stacked into a 4D group.

Since the birth of BM3D, there is no shortage of denoising methods originating from different disciplines. Interestingly, from the traditional Gaussian denoisers [14] with matrix and tensor representations to recently developed approaches using different deep neural network (DNN) architectures [15], nearly all the newly proposed methods claim to outperform the BM3D family [6], [11], [13], [16], [17]. However, some recent studies [18], [19] come to a different conclusion, and it is observed that many methods are normally verified based on a limited number (often less than three) of datasets, and the parameters of BM3D-based methods may not be fine-tuned with certain noise

---

• Zhaoming Kong and Lifang He are with the Computer Science and Engineering Department, Lehigh University, PA, 18015 USA.  
E-mail: zhk219@lehigh.edu

• Xiaowei Yang is with the school of Software Engineering, South China University of Technology, Guangdong Province, China.

estimation techniques [20], [21]. With a large number of existing methods [22] and outstanding survey papers [7], [8], [14], [15], [23], [24], [25], [26], [27], there still lacks a thorough comparison for the multi-dimensional image denoising tasks. In this paper, we intend to fill this gap by collecting and comparing different methods. The main contributions of this paper are three-fold:

1) We construct a new real-world indoor-outdoor color image (IOCI) and color video (IOCV) dataset with multiple different cameras. The dataset consist of images and video sequences captured based on both artificial indoor subjects and natural outdoor scenes under different lighting conditions, which are a good complement to current benchmark datasets with various challenging realistic backgrounds.

2) We compare over 60 multi-dimensional image denoising methods, and perform extensive experiments on both synthetic and real-world datasets. We adopt both subjective and objective metrics, and evaluate the denoising quality of compared methods with quantitative results and human visual ratings.

3) We make three interesting observations based on comprehensive experimental results. First the BM3D family still demonstrate very competitive performance for multiple denoising tasks in terms of both effectiveness and efficiency. Second, for traditional denoisers that learn solely on the noisy observation, we argue that a modified singular value decomposition (M-SVD) method is able to produce similar results with many tensor-based approaches. Furthermore, several DNN models trained with synthetic noise demonstrate impressive generalizability to real-world cases. Specifically, a fusion of collaborative and convolutional filtering (FCCF) model [28], and a U-Net [29] and ResNet [30] based denoising network (DRUNet) [31] produce the best results for the color image denoising task. The fast deep video denoising network (FastDVDNet) [32] and a nonlocal video denoising network (VNLNet) [33] show significant improvements for color video denoising compared with the benchmark video-BM4D (VBM4D) [17] implementation.

The rest of this paper is organized as follows. Section 2 introduces background knowledge, and Section 3 gives a brief review on related multi-dimensional image denoising techniques and datasets. Section 4 provides experimental results and discussions. Section 5 concludes the paper.

## 2 BACKGROUND

### 2.1 Symbols and Notations

In this paper, to represent multi-dimensional images, we mainly adopt the mathematical notations and preliminaries of tensors from [34]. Vectors and matrices are first- and second- order tensors which are denoted by boldface lowercase letters  $\mathbf{a}$  and capital letters  $\mathbf{A}$ , respectively. A higher order tensor (the tensor of order three or above) is denoted by calligraphic letters, e.g.,  $\mathcal{A}$ . An  $N$ th-order tensor is denoted as  $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_N}$ . The  $n$ -mode product of a tensor  $\mathcal{A}$  by a matrix  $\mathbf{U} \in \mathbb{R}^{P_n \times I_n}$ , denoted by  $\mathcal{A} \times_n \mathbf{U} \in \mathbb{R}^{I_1 \dots I_{n-1} P_n I_{n+1} \dots I_N}$  is also a tensor. The mode- $n$  matricization or unfolding of  $\mathcal{A}$ , denoted by  $\mathbf{A}_{(n)}$ , maps tensor elements  $(i_1, i_2, \dots, i_N)$  to matrix element  $(i_n, j)$  where  $j = 1 + \sum_{k=1, k \neq n}^N (i_k - 1) J_k$ , with  $J_k = \prod_{m=1, m \neq n}^{k-1} I_m$ . The Frobenius norm of a tensor  $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_N}$  is

defined as  $\|\mathcal{A}\|_F = \sqrt{\sum_{i_1=1} \dots \sum_{i_N=1} \mathcal{A}_{i_1 \dots i_N}^2}$ . The nuclear norm of a matrix is denoted by  $\|\mathbf{A}\|_*$ , which is the sum of all singular values of  $\mathbf{A}$ . The symbol of  $\otimes$  denotes the Kronecker product of matrices.

### 2.2 Noise modeling

Let us consider a noisy observation  $\mathcal{Y}$  and its underlying clean image  $\mathcal{X}$ , a general assumption of noise distribution is additive white Gaussian noise (AGWN) with variance  $\sigma^2$  represented by  $\mathcal{N}(0, \sigma^2)$ , and the degradation process is then given as

$$\mathcal{Y} = \mathcal{X} + \mathcal{N} \quad (1)$$

According to [35], noise modeling can be more complex and challenging since noise in real-world images may be multiplicative and signal dependent. Therefore, there are some non i.i.d Gaussian models designed for multi-dimensional images, such as the mixed Gaussian impulse noise of color images [36], [37], sparse random corruptions of video data [38], strip noise removal of MSI [39], and Rician noise reconstruction of MRI [40]. In this paper, our synthetic experiments and analysis are mainly based on the AGWN model because (i) a majority of compared methods could handle Gaussian noise, (ii) certain types of noise could be transformed to Gaussian distribution, and (iii) Romano et al. [41] recently point out that the removal of AGWN from an image is largely a solved problem, which may help explain the effectiveness of the simplified noise modeling in Eq. (1).

### 2.3 Nonlocal self-similarity

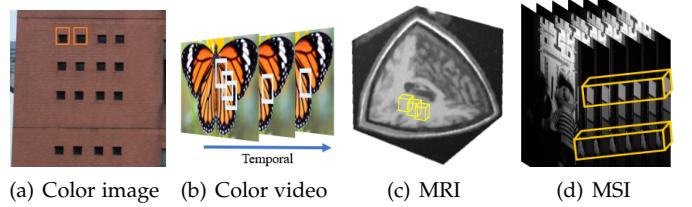


Fig. 1. The nonlocal self-similarity (NLSS) prior and patch representation of different multi-dimensional images.

The nonlocal self-similarity (NLSS) feature of natural images is probably the most important prior adopted by many different denoising methods. Briefly, NLSS refers to the fact that a local image patch often has many nonlocal similar patches to it across the image [42]. Usually, the similarity between two patches  $\mathcal{P}_A$  and  $\mathcal{P}_B$  is measured by their Euclidean distance  $d_{AB} = \|\mathcal{P}_A - \mathcal{P}_B\|$ . In practice, to save some time, the search for similar patches is restricted to a local window  $\Omega_{SR}$  with pre-defined size. As illustrated in Fig. 1, the patch representation and the rule of similar patch search (SPS) may vary for different types of multi-dimensional images. For example, SPS of color images can be conducted only on the luminance channel to reduce complexity; for video sequences, SPS is performed along both temporal and spatial dimensions; and for MRI and MSI data, a patch could be represented by a 3D cube or a square tube with multiple spectral bands.

TABLE 1

Related traditional denoisers for multi-dimensional image data with different noise modeling techniques and applications. 'CI': color image, 'CV': color video, 'MSI': multispectral imaging, 'HSI': hyperspectral imaging, 'MRI': magnetic resonance imaging.

Category	Representation	Methods	Noise Modeling	Applications	Key Words
Traditional denoisers	Matrix	AWT [43], [44], [45]	AWGN	CI, MSI, HSI	Adaptive wavelet thresholding methods
		DCT [46], [47]	AWGN	CI	Image denoising with discrete cosine transform (DCT)
		NLM [7], [48]	AWGN	CI	nonlocal means (NLM) algorithm with Gaussian kernels
		K-SVD [49], [50], [51]	AWGN	CI, MSI	Over-complete dictionary learning and sparse coding
		TCVD [52]	AWGN	CV	Combination of robust optical flow with NLM
		NLPICA [53], [54], [55]	AWGN	CI, MRI	Applications of SVD transform
		EPLL [56], [57]	AWGN	CI	A Bayesian method for whole image restoration
		LRMR [58]	AWGN	MRI	A low-rank matrix decomposition model
		GID [59]	AWGN	CI	External data guided and internal prior learning
		SPTWO [60]	AWGN	CV	NLPICA with optical flow estimation
		LSCD [61]	AWGN	CI, MSI, HSI	Spectral component decomposition with line feature
		MCWNMM [62], [63]	AWGN	CI	Extension of WNNM to color images
		TWSC [64]	AWGN	CI	A trilateral weighted sparse coding scheme
		VNLB [65]	AWGN	CV	A patch-based Bayesian model for video denoising
		NLH [66]	AWGN	CI	Nonlocal pixel similarity with Haar transform
		AMF [67]	Impulsive	CI	Impulsive noise removal by adaptive median filter
		NoiseClinic [68]	AWGN and Realistic	CI	A multiscale blind Bayes denoising algorithm
		LSM-NLR [37]	AWGN, Poisson and Impulsive	CV	Low-rank with laplacian scale mixture modeling
		FastHyde [69]	AWGN and Poisson	MSI, HSI	Low-rank and sparse representation
		NMoG [39]	AWGN, Stripe and Impulsive	MSI, HSI	Low-rank matrix model with non i.i.d Gaussian noise
		ODCT [70]	AWGN and Rician	MRI	A 3D DCT implementation
		ONLM [71]	AWGN and Rician	MRI	An optimized blockwise NLM Filter
		AONLM [72]	AWGN and Rician	MRI	Adaptive NLM with spatially varying noise levels
		RLMMSE [73]	Rician	MRI	A novel linear minimum mean square error estimator
		PRI-NLM [70]	AWGN and Rician	MRI	A rotationally invariant version of the NLM filter
		PRI-NLPICA [74]	AWGN and Rician	MRI	A two stage filter based on NLPICA and PRI-NLM
	Tensor	CBM3D [11]	AWGN	CI, CV	BM3D with opponent and YUV color mode transforms
		LRTA [75]	AWGN	MSI, HSI	A low Tucker rank tensor decomposition model
		PARAFAC [76]	AWGN	MSI	A rank-one candecomp/parafac (CP) model
		BM4D [13]	AWGN	MSI, HSI, MRI	Extention of BM3D to MSI and MRI using 3D patch
		VBM4D [17]	AWGN	CV	Extention of BM3D to color videos using 3D patch
		4DHOSVD [77], [78], [79]	AWGN	CI, MSI, HSI, MRI	Applications of 4DHOSVD transform
		TDL [80]	AWGN	MSI, HSI	A Tucker based tensor dictionary learning algorithm
		KTSDV [81]	AWGN	MSI, HSI	A t-SVD based tensor dictionary learning algorithm
		LRTV [82]	AWGN	MSI, HSI	Low-rank model with total variation (TV) regularization
		ITSReg [83]	AWGN	MSI, HSI	Intrinsic Tensor Sparsity Regularization
		LLRT [84]	AWGN	CI, MSI	low-rank tensor with hyper-laplacian regularization
		LLRGTV [85]	AWGN	MSI, HSI	Low Tucker rank decomposition with total variation
		WTR1 [86]	AWGN	CI	A weighted rank-one CP decomposition model
		ILR-HOSVD [87]	AWGN	MRI	A recursive low Tucker rank model with rank estimation
		MSr-SVD [19], [88]	AWGN	CI, MRI, MSI, HSI	An efficient one-step t-SVD implementation
		NCMeet [89]	AWGN	MSI, HSI	Low-rank tensor model with iterative regularization
		LTDL [90]	AWGN	MSI, HSI	A low-rank tensor dictionary learning method
		NLTA-LSM [91]	AWGN and Poisson	MSI, HSI, MRI	Low Tucker rank with laplacian scale mixture modeling

### 3 RELATED WORKS

In this section, we briefly introduce related multi-dimensional denoising methods and datasets, which are summarized in Table 1, Table 2 and Table 3, respectively. The timeline of representative denoising methods and datasets is illustrated in Fig. 4. In our paper, the denoising methods are roughly divided into two categories, namely 'traditional denoisers' and 'DNN methods', depending on whether neural network architectures are used. More detailed description of related works are available in [7], [8], [14], [15], [23], [24], [25], [26], [27].

#### 3.1 Traditional denoisers

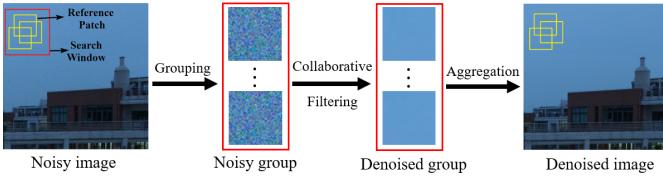


Fig. 2. Illustration of the grouping-collaborative filtering-aggregation framework for traditional denoisers.

For traditional denoisers, training and denoising are usually accomplished only with the noisy image. To utilize the NLSS prior, the most popular and successful framework attributed to BM3D [6] mainly follows three consecutive stages: grouping, collaborative filtering and aggregation. The flowchart of this effective three-stage paradigm is illustrated in Fig. 2.

#### 3.1.1 Grouping

For every  $d$ -dimensional noisy image patch  $\mathcal{P}_n$ , based on certain patch matching criteria [47], [60], [92], [93], [94], [95], the grouping step stacks  $K$  similar (overlapping) patches located within a local window  $\Omega_{SR}$  into a  $d+1$ -dimensional group. For example, consider a 3D patch  $\mathcal{P}_n \in \mathbb{R}^{H \times W \times N}$ , where  $H$ ,  $W$  and  $N$  represents height, width and the number of channels or spectral bands, respectively, the 4D group of  $K$  patches can be directly represented by a fourth-order tensor  $\mathcal{G}_n \in \mathbb{R}^{H \times W \times N \times K}$ , or a 2D matrix  $\mathbf{G}_n \in \mathbb{R}^{HWN \times K}$  if every patch  $\mathcal{P}_n$  is reshaped into a long vector  $\mathbf{p}_n \in \mathbb{R}^{HWN}$ .

#### 3.1.2 Collaborative filtering

Collaborative filters operate on the noisy patch group  $\mathcal{G}_n$  to estimate the corresponding underlying clean group  $\mathcal{G}_c$  via

$$\hat{\mathcal{G}}_c = \arg \min_{\mathcal{G}_c} \|\mathcal{G}_n - \mathcal{G}_c\|_F^2 + \rho \cdot \Psi(\mathcal{G}_c) \quad (2)$$

or in the matrix form

$$\hat{\mathbf{G}}_c = \arg \min_{\mathbf{G}_c} \|\mathbf{G}_n - \mathbf{G}_c\|_F^2 + \rho \cdot \Psi(\mathbf{G}_c) \quad (3)$$

where  $\|\mathcal{G}_n - \mathcal{G}_c\|_F^2$  or  $\|\mathbf{G}_n - \mathbf{G}_c\|_F^2$  indicates the conformity between the clean and noisy groups, and  $\Psi(\cdot)$  is a regularization term for certain priors. For example, to model the nonlocal redundancies, the low-rank prior is adopted in [54], [63], [84], [91] with  $\Psi(\mathbf{G}_c) = \|\mathbf{G}_c\|_*$  for matrix and  $\Psi(\mathcal{G}_c) = \sum_{n=1}^N a_{(n)} \|\mathcal{G}_{c(n)}\|_*$  for tensor [96]. In addition,

the dictionary learning model with over-complete representations [49], [50], [64] is utilized to reconstruct  $\mathbf{G}_c$  with a dictionary  $\mathbf{D}$  and sparse coding coefficients  $\mathbf{C}$  via

$$\hat{\mathbf{C}} = \arg \min_{\mathbf{C}} \|\mathbf{G}_n - \mathbf{DC}\|_F^2 + \lambda \|\mathbf{C}\|_1 \quad (4)$$

where  $\lambda \|\cdot\|$  is the regularization term that enforces sparsity constraint on  $\mathbf{C}$ . Once  $\hat{\mathbf{C}}$  is computed, the latent clean patch group  $\hat{\mathbf{G}}_c$  can be estimated as  $\hat{\mathbf{G}}_c = \hat{\mathbf{D}}\hat{\mathbf{C}}$ . In [80] and [81], Eq. (4) is extended to tensor for MSI denoising with higher-order SVD (HOSVD) [97], [98] and tensor-SVD (t-SVD) [99], [100] transforms, respectively. A simple and effective method is to model the sparsity with certain thresholding techniques [43], [101] to attenuate the noise. For example, the hard-thresholding technique is adopted by the BM3D family and some state-of-the-art tensor-based methods [19], [77], which attempts to shrink the coefficients  $\mathcal{T}(\mathcal{G}_n)$  in the transform-domain [8] under a threshold  $\tau$  via

$$\mathcal{G}_t = \begin{cases} \mathcal{T}(\mathcal{G}_n), & |\mathcal{T}(\mathcal{G}_n)| \geq \tau \\ 0, & |\mathcal{T}(\mathcal{G}_n)| < \tau \end{cases} \quad (5)$$

where  $\mathcal{T}$  represents a pre-defined or learned transform. The estimated clean group  $\hat{\mathcal{G}}_c$  is obtained by inverting the transform via

$$\hat{\mathcal{G}}_c = \mathcal{T}^{-1}(\mathcal{G}_t) \quad (6)$$

The popularity of SVD-based transforms in traditional denoisers largely results from its invertible orthogonal bases.

### 3.1.3 Aggregation

To further smooth out noise, the estimated clean patches of  $\hat{\mathcal{G}}_c$  are averagely written back to their original location. More specifically, at the pixel level, every pixel  $\hat{p}_i$  of the denoised image is the (weighted) average of all pixels at the same position of filtered group  $\hat{\mathcal{G}}_c$ , which can be formulated as

$$\hat{p}_i = \sum_{\hat{p}_{i_k} \in \hat{\mathcal{G}}_c} w_{i_k} \hat{p}_{i_k} \quad (7)$$

where  $w_{i_k}$  and  $\hat{p}_{i_k}$  denote weight and local pixel of  $\hat{\mathcal{G}}_c$ , respectively.

The major difference among various traditional denoisers mainly derives from the collaborative filtering scheme, which is often decided by the choice of transforms and algebraic representation. Intuitively, reshaping the 4D group  $\mathcal{G}$  of (2) into the 2D matrix  $\mathbf{G}$  of (3) may break the internal structure of multi-dimensional images. Therefore, a typical assumption [19], [77], [78], [80], [81], [84], [90], [102] is that tensor representation and decomposition techniques could help preserve more structure information, based on the fact that multi-dimensional images could be naturally represented by multi-dimensional array. However, the conventional and widely-used tensor model may fall into the unbalance trap [103], [104], leading to unsatisfactory image restoration performance. In this paper, we further show that with slight modifications, a simple modified SVD (MSVD) implementation could produce competitive results for color image denoising compared with tensor-based methods. More details are given in Appendices.

## 3.2 DNN methods

The most recent development of image denoising is mainly brought by the applications of deep neural networks (DNNs), which demonstrate excellent performance in the latest denoising competition [140] based on the smartphone image denoising dataset (SIDD) [141]. Different from traditional denoisers that use only internal information of the noisy observation, DNN methods usually adopt the supervised training strategy guided by external priors and datasets to minimize the distance  $\mathcal{L}$  between predicted and clean images via

$$\min_{\theta} \sum_i \mathcal{L}(\mathcal{F}_{\theta}(\mathcal{Y}_i), \mathcal{X}_i) + \rho \cdot \Psi(\mathcal{F}_{\theta}(\mathcal{Y}_i)) \quad (8)$$

where  $\mathcal{X}_i$  and  $\mathcal{Y}_i$  are clean/noisy image (patch) pairs,  $\mathcal{F}_{\theta}$  with parameter  $\theta$  is a nonlinear function that maps noisy patches onto predicted clean patches, and  $\Psi(\cdot)$  represents certain regularizers [109], [142]. Early methods [143], [144] work with the known shift blur function and weighting factors, and Burger et al. [105] show that a plain multi-layer perceptron (MLP) network is able to compete with BM3D at certain noise levels. To extract latent features and exploit the NLSS prior, a more commonly used model is the convolutional neural network (CNN) [145], which is suitable for multi-dimensional data processing [146], [147], [148], [149] with flexible size of convolution filters and local receptive fields. Convolution operations are first applied to image denoising in [150], and Fig. 3 illustrates a simple CNN denoising framework with three convolution layers.

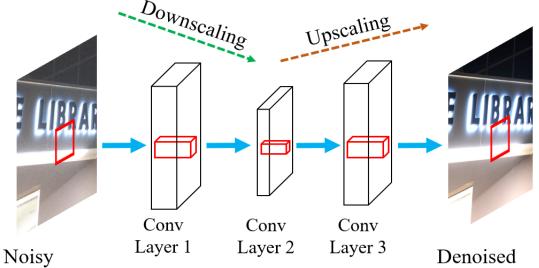


Fig. 3. Illustration of a simple CNN denoising framework with three convolution layers.

Due to the simplicity and effectiveness of CNN, it is widely adopted by different DNN denoising algorithms, and the variations of CNN-based networks are quite extensive. For example, for color images, the denoising CNN (DnCNN) [107] incorporates batch normalization (BN) [151], rectified linear unit (ReLU) [152] and residual learning [30] into the CNN model. A generative adversarial network (GAN) blind denoiser (GCBD) [136] introduces GAN [153] to resolve the problem of unpaired noisy images. Recently, a graph convolution denoising network (GCDN) is presented in [135] to capture self-similar information. For color videos, VNLNet [33] proposes a new approach to combine video self-similarities with CNN. FastDVDNet [32] achieves real-time video denoising without a costly motion compensation stage. For MSI, HSID-CNN [112] and HSI-sDeCNN [113] assign both spatial and spectral information of MRI data simultaneously to CNN. A 3D Quasi-Recurrent Neural Network (QRNN3D) [131] utilizes 3D convolutions to extract structural spatio-spectral correlation of MSI data, and

TABLE 2

Related DNN denoising methods for multi-dimensional image data with different noise modeling techniques and applications. 'CI': color image, 'CV': color video, 'MSI': multispectral imaging, 'HSI': hyperspectral imaging, 'MRI': magnetic resonance imaging.

Category	Network Architectures	Methods	Noise Modeling	Applications	Key Words
DNN methods	CNN/NN	MLP [105]	AWGN	CI	A multilayer perceptron model
		TNRD [106]	AWGN	CI	A trainable nonlinear diffusion model
		DnCNN [107], [108]	AWGN	CI, MRI	CNN with Residual learning and batch normalization
		NLNet [109]	AWGN	CI	A nonlocal CNN model
		UDNet [110]	AWGN	CI	A robust and flexible CNN model
		FFDNet [111]	AWGN	CI	A flexible model with tunable input noise levels
		HSID-CNN [112]	AWGN	MSI, HSI	A 2D and 3D combined CNN model
		HSI-SDeCNN [113]	AWGN	MSI, HSI	A single CNN model with spatial and spectral information
		VNLNet [33]	AWGN	CV	The first nonlocal CNN model for video
		ViDeIN [114]	AWGN	CV	Combination of spatial and temporal filtering with CNN
		SADNet [115]	AWGN	CI	A novel spatial-adaptive CNN model
		FastDVINet [32]	AWGN	CV	Real-time video denoising without flow estimation
		LIDIA [116]	AWGN	CI	A lightweight denoiser with instance adaptation
		DRUNet [31]	AWGN	CI	A plug-and-play method with deep denoiser prior
		ADRN [117]	AWGN	MSI, HSI	A deep residual network model with channel attention scheme
		HSI-DeNet [118]	AWGN and Strip	MSI, HSI	A CNN method with mixed noise modeling
		Noise2noise [119]	AWGN and Poisson	CI	A CNN model trained only with noisy data.
		VDNet [120]	AWGN and Realistic	CI	A variational inference method with noise estimation
		CBNet [121]	Realistic	CI	A convolutional blind denoising network
		PRI-PB-CNN [122]	AWGN and Rician	MRI	Combination of sliding window scheme and 3D CNN
		MIFCN [123]	Realistic	MRI	A fully convolutional network model
		DIDN [124]	AWGN and Realistic	CI	A deep iterative down-up CNN model
		DRDN [125]	Realistic	CI	A dynamic residual dense network model
		SGN [126]	Realistic	CI	A self-guided network with top-down architecture
		FCCF [28]	AWGN and Realistic	CI	Fusion of collaborative filtering and CNN
		DBF [127]	AWGN and Realistic	CI	Integration of CNN into a boosting algorithm
		RIDNet [128]	AWGN and Realistic	CI	A single stage model with feature attention
		AINDNet [129]	AWGN and Realistic	CI	Transfer learning with adaptive instance normalization
		Self2Self [130]	AWGN and Realistic	CI	A self-supervised model trained only with noisy images
		QRNN3D [131]	AWGN and Stripe	MSI, HSI	A deep recurrent neural network model with 3D convolution
		CycleISP [132]	Realistic	CI	A camera pipeline model in forward and reverse directions
		DANet [133]	Realistic	CI	A Bayesian framework for noise removal and generation.
		MIRNet [134]	Realistic	CI	A multi-scale model with parallel convolution streams
	CNN + GAN	GCDN [135]	AWGN and Realistic	CI	A graph convolution denoising network
		GCBNet [136]	AWGN and Realistic	CI	A GAN-based blind denoiser
		SCGAN [137]	AWGN	CI	Unsupervised modeling with self-consistent GAN
		DnGAN [138]	AWGN and Blur	CI	GAN with maximum a posteriori (MAP) framework
		ADGAN [139]	Realistic	CI	Attentive GAN with noise domain adaptation

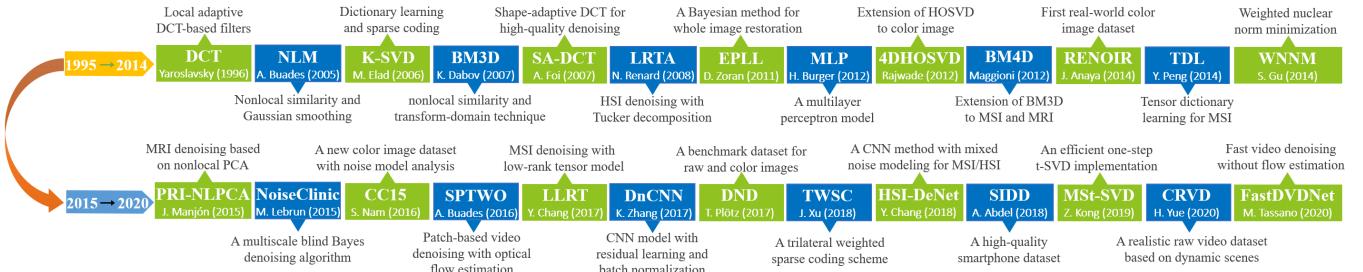


Fig. 4. Timeline of representative multi-dimensional image denoising methods and datasets.

employ a quasi-recurrent pooling function to capture the global correlation along spectrum. For MRI denoising, a multi-channel DnCNN (MCDnCNN) [108] network extends DnCNN to handle 3D volumetric data, and the prefiltered rotational invariant patch-based CNN (PRI-PB-CNN) [122] trains the CNN model with MRI data pre-filtered by PRI nonlocal principal component analysis (PRI-NLPCA) [74].

Despite the effectiveness of DNN methods, they could be two-folded weapons, which enjoy three major advantages and also face the same challenges. First, DNN methods are able to utilize external information to guide the training process, and thus may not be restricted to the theoretical and practical bound of traditional denoisers [154]. However, DNN methods rely heavily on the quality of the training datasets, and also certain prior information such as ISO, shutter speed and camera brands, which are not always available in practice. Second, DNN methods could take advantage of the advanced GPU devices for acceleration and achieve real-time denoising [32], [155] for certain tasks. But the expensive computational resources may not be

accessible to ordinary users and researchers. Third, the deep, flexible and sophisticated structure of DNN methods is capable of extracting the latent features underlying the noisy images, but compared with the implementation of CBM3D that only needs to store four small pre-defined transform matrices, the complex networks with millions of parameters may drastically increase the storage cost.

### 3.3 Datasets

In this section, we briefly introduce popular multi-dimensional image datasets used for synthetic and real-world experiments. The information is summarized in Table 3, with more details given in [27], [28], [141], [156], [157], [158], [159], [160], [161], [162], [163], [164], [165]. Some sample images of the datasets are illustrated in Fig. 5. Usually, datasets used for synthetic experiments consist of noise-free images acquired under ideal conditions with sufficient light and careful camera settings, whereas real-world images are inevitably contaminated by noise to various degrees, decided by the environments and equipments.

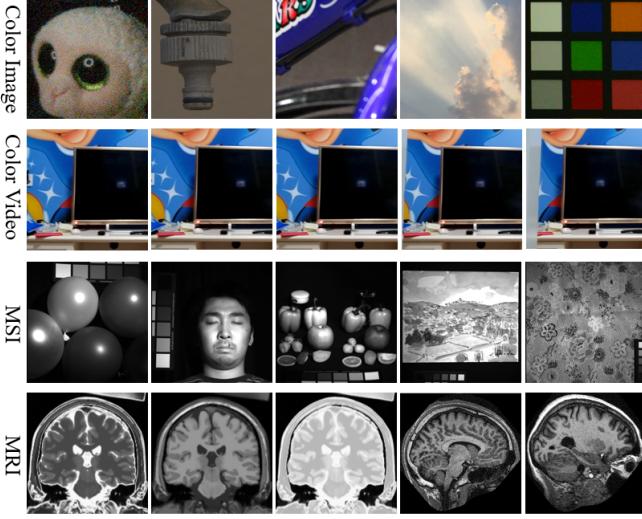


Fig. 5. Illustration of multi-dimensional image datasets. From the first row to the fourth row: color image, color video, MSI and MRI data.

TABLE 3

Information of popular multi-dimensional image denoising datasets for synthetic and real-world experiments. 'GT': ground-truth, 'A': Available, 'N/A': Not-Available, 'F': number of frames.

Type	Name	Experiments	GT	# Cameras	Image size	# Images
Color image	Kodak [166]	Synthetic	A	/	768 × 512 × 3	24
	RENOIR [157]	Real-world	A	3	3684 × 2760 × 3	120
	Nam-CC15 [156]	Real-world	A	3	512 × 512 × 3	15
	Nam-CC60 [156]	Real-world	A	3	500 × 500 × 3	60
	PolyU [158]	Real-world	A	5	512 × 512 × 3	100
	DND [18]	Real-world	N/A	4	512 × 512 × 3	1000
	SIDD [141]	Real-world	A	5	256 × 256 × 3	1280
	High-ISO [28]	Real-world	A	2	512 × 512 × 3	90
Color video	Our IOCI	Real-world	A	9	1024 × 1024 × 3	409
	Set8 [167]	Synthetic	A	/	960 × 540 × 3 × F	4
	DAVIS [159]	Synthetic	A	/	854 × 480 × 3 × F	30
	CRVD [160]	Real-world	A	1	1920 × 1080 × 3 × F	61
MSI	Our IOCIV	Real-world	A	4	512 × 512 × 3 × F	39
	CAVE [161]	Synthetic	A	1	512 × 512 × 31	32
	ICVL [162]	Synthetic	A	1	1392 × 1300 × 31	201
MRI	HHD [163]	Real-world	N/A	1	1392 × 1040 × 31	77
	BrainWeb [164]	Synthetic	A	/	181 × 217 × 181	3
	OASIS [165]	Real-world	N/A	/	256 × 256 × 128	2

### 3.3.1 Color image and video datasets

Some recent works [168] target raw image denoising based on the characteristics of photosensors, but the raw data are not always accessible, so we focus on the standard RGB (sRGB) color space. Compared to other types of images, the abundance of real-world color image datasets largely results from the relatively low cost of generating clean reference images from noisy observations. Briefly, a real low-light image noise reduction dataset (RENOIR) [157] is the first attempt to obtain low-noise reference images with low light sensitivity (ISO = 100) and long exposure time. The Darmstadt noise dataset (DND) [18] utilizes a more careful post-processing technique to produce higher quality reference images based on low-ISO images. A more common strategy adopted by a variety of datasets [27], [28], [141], [158], [160] is called 'image averaging', which captures the same and unchanged scene for many times and computes their mean value to obtain the corresponding noise-free image. The rationale of this simple strategy is that for each

pixel, the random noise is assumed to be larger or smaller than 0, and thus can be greatly suppressed by sampling the same pixel for many times [158].

Compared to the passion for collecting real-world color image datasets, fewer efforts have been made to produce realistic clean color videos. The image averaging strategy can not be directly applied to videos, and the difficulty lies mainly in continuously capturing noisy-clean video pairs for dynamic scenes. Besides, the strategy of generating clean dynamic videos using low ISO and high exposure time may cause motion blur. Recently, the captured raw video denoising (CRVD) dataset [160] is introduced based on a wise frame-by-frame strategy. The authors propose to capture raw video frames by moving controllable static objects to manually create motions for them. For each motion,  $M$  raw noisy images are captured and their mean image is regarded as the noise-free video frame. Finally, all the noise-free frames are grouped together according to their temporal order to generate the clean video.

### 3.3.2 MSI and MRI datasets

In this subsection, we briefly introduce the MSI and MRI datasets listed in Table 3. Briefly, the emergence of MSI and MRI reflects the development of new imaging techniques, but compared to color images, collecting MSI and MRI data are of greater difficulty and also more expensive. To the best of our knowledge, there is no publicly available realistic MSI and MRI dataset with noisy/clean pairs.

MSI cameras are mostly spectroscopic devices, which need to be carefully calibrated in order to obtain a reliable spectral information [169]. The representative CAVE dataset [161] is constructed with a generalized assorted pixel (GAP) camera, it includes 32 scenes of a wide variety of real-world materials and objects. Each image has the size of  $512 \times 512$  in space and includes full spectral resolution reflectance data ranging from 400 nm to 700 nm at 10 nm steps, which leads to 31 bands. The ICVL dataset [162] is acquired using a Specim PS Kappa DX4 camera, and it includes more than 200 different images collected at  $1392 \times 1300$  spatial resolution, and the data are downsampled to 31 spectral channels from 400nm to 700nm at 10nm increments. The real-world harvard hyperspectral dataset (HHD) dataset [163] includes 75 images of size  $1040 \times 1392 \times 31$  collected under daylight, artificial and mixed illumination. The images were captured using a commercial hyperspectral camera (Nuance FX, CRI Inc) with 31 wavelength bands ranging from 420nm to 720nm at 10nm steps.

The MRI technique uses a strong magnetic field and radio waves to create detailed images of the organs and tissues within the body, and it could be regarded as an extension of white light imaging to incorporate better spectral resolution [170]. The acquisition of MRI data is often expensive and time-consuming. For synthetic experiments, the Brainweb dataset contains a set of MRI data volumes produced by an MRI simulator, which are often used as 'ground-truth' MR images for different validation purpose. Full 3D data volumes have been simulated using three sequences (T1-, T2-, and proton-density- (PD-) weighted) and a variety of slice thicknesses, noise levels, and levels of intensity non-uniformity. For realistic OASIS brain image datasets, the selected 3D T1-weighted (T1w) data of size  $256 \times 256 \times 128$

are acquired by an MP-RAGE volumetric sequence on a Siemens 1.5T Vision scanner, with repetition time (TR) 9.7 msec, echo time (TE) 4.0 msec, flip angle 10 degrees, inversion time (TI) 10 msec, duration time of 200 ms, and voxel resolution of  $1 \times 1 \times 1.25$  mm<sup>3</sup>.

### 3.3.3 The proposed color image and video datasets

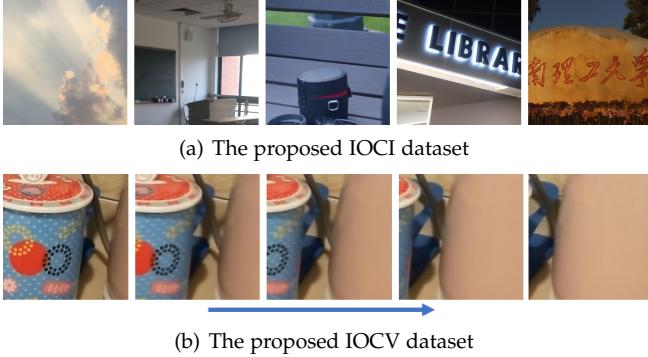


Fig. 6. Illustration of the proposed IOCI and IOCV dataset.

In this subsection, we briefly introduce the motivation and details regarding the setup and protocol followed by our indoor-outdoor color image (IOCI) and video (IOCV) datasets, some examples are shown in Fig. 6.

**IOCI.** Many existing datasets are mainly restricted to static indoor scenes where the lighting conditions can be manually controlled to simulate different illumination. But in a lot of real-world cases, photos are taken in outdoor environments where the objects and source of light may be more complicated. In our IOCI dataset, nine different cameras<sup>1</sup> are used to collect images of both indoor and outdoor scenes. Unlike previous works [28], [141], [158] that use pre-defined camera settings such as ISO, shutter speed and aperture, we mostly adopt the settings of the cameras' 'auto mode', which minimizes human interference, and is also suitable for the uncontrollable outdoor environments. Captured images that display obvious misalignment and illumination differences are discarded. Each indoor and outdoor 'ground-truth' image is obtained by averaging at least 100 and 30 images of the same scene, respectively.

**IOCV.** The carefully-designed frame-by-frame strategy of the CRVD dataset is able to produce high-quality clean reference videos, but it also requires a lot of human effort, and the manually-created motions may not be continuous. Therefore, we adopt a video-by-video strategy, and instead of manually moving controllable objects, we propose to move cameras automatically. The procedure of generating mean videos as ground-truth is illustrated in Fig. 7. Specifically, we choose light-weight cameras to avoid clear shaking and misalignment, and fix the cameras to a rotatable tripod ball-head placed on top of a professional motorized slider. The slider and the tripod ball-head could be set to move and rotate at different speeds, which simulate the movement of observed objects in more than one directions. Both the slider and cameras are controlled remotely to avoid human interference. Noisy sequences are captured repeatedly for at least 30 times to generate the corresponding 'ground-truth' reference video by their mean value.

<sup>1</sup>Six cameras are used in previous experiments [19]

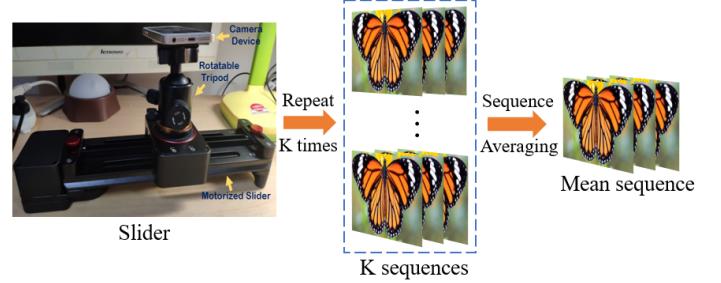


Fig. 7. The procedure of generating mean video sequences with a motorized slider. The camera is fixed to a rotatable tripod ball-head placed on top of the slider.

## 4 EXPERIMENTS

In this section, we compare the performance of different methods for the denoising task of color images, color videos, MSI and MRI data. All implementations are provided by the authors or downloaded from the authors' personal websites. The BM3D family normally consist of two stages, where the first stage utilizes hard-thresholding to provide an initial estimate, which is then smoothed with a nonlocal empirical Wiener filter to produce a final filtered image. In our experiments, we refer to these two output stages as 'BMXD1' and 'BMXD2', respectively. For DNN methods that require GPU devices, we use Google Colab's free GPU. All other experiments are performed on a personal computer equipped with Core(TM) i5-7500 @ 3.4 GHz and 16GB RAM.

### 4.1 Results for Real-World Color Image Datasets

#### 4.1.1 Experimental Settings

We compare the performance of over 25 effective methods on four real-world color image datasets, namely CC15, C-C60, PolyU and our IOCI. Typically, the decisive parameters for traditional denoisers are input noise level, patch size and the number of local similar patches, etc, whereas the number of layers, learning rate and weight size are essential for DNN methods. It is noticed that many of the compared methods are designed for synthetic datasets, while ideally, in real-world cases, the parameters of all compared methods should be fine tuned to produce the best possible performance, but this could be computationally expensive and also the clean reference images for training and validation are always not available. Therefore, a more practical way is to selectively apply effective pre-trained models. Since most of the DNN implementations offer two to six pre-trained models with different parameter settings for testing, for fair comparison, we tune the input noise level for each Gaussian denoiser from four different values, which could be regarded as four pre-trained models corresponding to 'low', 'medium-low', 'medium-high' and 'high' denoising modes. We report the best average values of all compared methods on each dataset. Also, similar to [19], the best result of CBM3D on every image is included to further investigate its effectiveness, and this implementation is termed 'CB-M3D\_best'. Peak signal-to-noise ratio (PSNR) and structural similarity (SSIM) [171] are employed for objective evaluations. Normally, the higher the PSNR and SSIM values, the better the quality of denoised images.

TABLE 4

Average PSNR, SSIM values and computational time (second) on four realistic color image datasets. The average time is calculated based on the PolyU dataset, and 'N/A' means the models can not handle certain size of images. The best results (excluding CBM3D\_best) are black bolded.

Method	Dataset	CC15	CC60	PolyU	IOCI									Implementation	Time (s)	
		# Images	15	60	100	55	25	30	36	67	36	50	39	FUJIFILM X100T		
		/	Noisy	33.41 0.8476	34.51 0.8712	35.97 0.9325	38.94 0.9482	39.17 0.9379	36.94 0.9185	38.21 0.9359	42.89 0.9753	34.49 0.9251	37.84 0.9356	39.39 0.9307	/	/
Traditional denoisers	LCSD	36.20 0.9359	37.69 0.9226	37.97 0.9641	40.65 0.9732	41.45 0.9793	39.22 0.9633	39.84 0.9669	39.62 0.9627	39.53 0.9625	44.32 0.9663	35.18 0.9560	38.69 0.9669	38.75 0.9440	MATLAB + MEX	6.45
	LLRT	37.36 0.9559	38.51 0.9636	38.51 0.9707	41.84 0.9784	42.53 0.9816	39.54 0.9669	40.02 0.9676	39.72 0.9663	45.71 0.9899	34.93 0.9557	36.63 0.9687	39.19 0.9633	39.19 0.9633	MATLAB	>1000
	GID	37.02 0.9459	38.41 0.9633	38.37 0.9673	40.86 0.9743	41.60 0.9790	39.52 0.9653	40.12 0.9642	40.16 0.9670	44.94 0.9887	36.11 0.9599	40.50 0.9729	42.31 0.9750	42.31 0.9750	MATLAB	38.88
	MCWNNM	37.72 0.9568	39.03 0.9698	38.51 0.9671	41.47 0.9774	42.07 0.9795	39.46 0.9610	39.87 0.9567	40.18 0.9628	45.37 0.9894	35.84 0.9516	40.71 0.9734	42.48 0.9757	42.48 0.9757	MATLAB	228.16
	NLH	38.49 0.9647	39.86 0.9759	38.36 0.9655	41.77 0.9790	42.72 0.9841	38.84 0.9593	40.44 0.9644	39.94 0.9630	46.00 0.9905	35.73 0.9552	40.69 0.9740	42.80 0.9776	42.80 0.9776	MATLAB + MEX	51.86
	TWSI	37.90 0.9592	39.66 0.9759	38.62 0.9674	41.65 0.9767	42.52 0.9824	39.71 0.9651	40.27 0.9617	40.12 0.9619	45.48 0.9896	35.41 0.9385	40.71 0.9720	42.22 0.9728	42.22 0.9728	MATLAB	368.68
	M-SVD	37.62 0.9542	39.27 0.9705	38.56 0.9671	41.81 0.9788	42.54 0.9823	39.94 0.9671	40.66 0.9660	40.35 0.9690	45.69 0.9897	36.24 0.9606	40.69 0.9735	42.52 0.9750	42.52 0.9750	MATLAB	75.88
	4DHOSVD	37.51 0.9559	39.15 0.9729	38.51 0.9673	41.41 0.9771	42.14 0.9797	39.82 0.9658	40.68 0.9664	40.36 0.9671	45.56 0.9901	36.17 0.9600	40.69 0.9736	42.47 0.9753	42.47 0.9753	MATLAB	80.69
	CMSt-SVD	37.95 0.9588	39.75 0.9756	38.82 0.9694	41.99 0.9794	42.75 0.9840	40.08 0.9674	40.84 0.9668	40.53 0.9674	45.89 0.9903	36.40 0.9616	40.85 0.9743	42.68 0.9770	42.68 0.9770	MATLAB + MEX	6.18
	CBM3D1	37.56 0.9562	39.21 0.9729	38.52 0.9676	41.71 0.9780	42.50 0.9827	39.93 0.9660	40.56 0.9666	40.45 0.9681	45.56 0.9897	36.30 0.9603	40.62 0.9725	42.53 0.9762	42.53 0.9762	MATLAB + MEX	0.95
	CBM3D2	37.70 0.9572	39.41 0.9740	38.69 0.9695	41.69 0.9780	42.54 0.9836	39.97 0.9669	40.77 0.9668	40.55 0.9693	45.70 0.9902	36.38 0.9610	40.75 0.9737	42.65 0.9767	42.65 0.9767	MATLAB + MEX	4.38
	CBM3D_best	37.95 0.9621	39.68 0.9775	38.81 0.9712	42.08 0.9808	42.89 0.9851	40.48 0.9741	41.25 0.9758	41.16 0.9783	45.81 0.9904	36.66 0.9640	40.89 0.9742	43.14 0.9805	43.14 0.9805	MATLAB + MEX	/
DNN methods	DnCNN	37.47 0.9557	39.32 0.9742	38.51 0.9663	40.81 0.9717	41.91 0.9790	39.92 0.9662	39.34 0.9639	40.22 0.9640	43.94 0.9599	36.30 0.9726	40.26 0.9690	41.57 0.9736	41.57 0.9736	MATLAB + MEX	3.28
	FFDNet	37.68 0.9563	39.73 0.9770	38.56 0.9658	41.67 0.9768	42.55 0.9824	40.05 0.9645	40.49 0.9707	40.49 0.9645	45.71 0.9901	36.47 0.9607	40.75 0.9733	42.44 0.9759	42.44 0.9759	MATLAB + MEX	1.99
	CBDNet	36.36 0.9294	37.46 0.9482	36.93 0.9457	39.43 0.9729	41.26 0.9788	37.11 0.9505	38.56 0.9586	37.62 0.9428	42.17 0.9819	35.09 0.9493	37.68 0.9601	40.21 0.9621	40.21 0.9621	Python + GPU	3.58
	DBF	38.38 0.9644	40.59 0.9811	38.87 0.9698	41.36 0.9765	42.87 0.9841	39.98 0.9669	39.89 0.9654	40.26 0.9660	44.42 0.9804	36.45 0.9631	40.67 0.9734	42.36 0.9751	42.36 0.9751	Python + GPU	0.32
	DIDN	36.04 0.9456	N/A	37.32 0.9523	40.66 0.9739	41.58 0.9760	38.09 0.9500	38.22 0.9535	38.50 0.9493	43.89 0.9840	35.25 0.9485	39.66 0.9647	41.17 0.9655	41.17 0.9655	Python + GPU	0.18
	FCCF	39.04 0.9675	40.89 0.9819	38.85 0.9794	41.91 0.9852	43.07 0.9856	40.11 0.9656	40.39 0.9659	40.65 0.9680	45.38 0.9888	36.66 0.9640	40.93 0.9747	42.78 0.9782	42.78 0.9782	MATLAB + MEX	36.36
	NLNet	37.56 0.9554	39.32 0.9744	38.58 0.9684	41.43 0.9759	42.18 0.9811	39.96 0.9687	40.06 0.9637	40.61 0.9709	44.84 0.9875	36.34 0.9617	40.63 0.9729	42.19 0.9739	42.19 0.9739	MATLAB + MEX	49.98
	UDnet	36.95 0.9425	38.96 0.9708	38.05 0.9683	40.40 0.9774	41.36 0.9631	39.35 0.9591	38.94 0.9683	43.15 0.9811	43.15 0.9582	36.09 0.9660	39.51 0.9675	41.12 0.9675	41.12 0.9675	MATLAB + MEX	5.69
	VDNNet	35.86 0.9454	N/A	37.58 0.9574	40.97 0.9763	42.37 0.9831	38.23 0.9555	37.88 0.9562	38.60 0.9556	44.27 0.9844	35.49 0.9530	39.93 0.9699	41.19 0.9701	41.19 0.9701	Python + GPU	0.21
	AINDNet	35.88 0.9113	36.98 0.9305	37.22 0.9460	38.28 0.9748	39.21 0.9759	36.53 0.9544	36.68 0.9581	38.83 0.9527	40.18 0.9812	34.49 0.9523	37.20 0.9676	38.50 0.9656	38.50 0.9656	Python + GPU	1.29
	CycleISP	35.40 0.9158	36.87 0.9393	37.61 0.9546	41.24 0.9766	41.79 0.9770	38.43 0.9515	38.69 0.9605	38.60 0.9518	44.71 0.9874	35.50 0.9489	39.85 0.9682	41.58 0.9688	41.58 0.9688	Python + GPU	0.38
	RIDNet	36.83 0.9406	38.11 0.9609	38.57 0.9702	41.08 0.9770	42.18 0.9801	38.88 0.9621	38.69 0.9588	39.02 0.9587	44.40 0.9873	35.75 0.9570	40.08 0.9719	41.58 0.9750	41.58 0.9750	Python + GPU	1.06
	SADNet	37.80 0.9521	N/A	39.30 0.9725	40.74 0.9798	42.03 0.9797	39.55 0.9646	39.91 0.9656	39.72 0.9626	44.69 0.9886	36.38 0.9620	40.72 0.9740	41.91 0.9752	41.91 0.9752	Python + GPU	0.33
	DANet	37.17 0.9526	38.63 0.9694	41.76 0.9771	41.59 0.9788	42.47 0.9833	38.46 0.9545	39.39 0.9661	38.80 0.9527	44.71 0.9874	36.21 0.9608	39.92 0.9673	41.44 0.9693	41.44 0.9693	Python + GPU	0.12
	Self2Self	36.26 0.9466	37.88 0.9672	37.66 0.9563	N/A	Python + GPU	>2000									
	MIRNet	36.06 0.9416	37.25 0.9556	37.49 0.9556	40.72 0.9779	41.68 0.9777	38.28 0.9539	38.60 0.9591	38.55 0.9522	43.66 0.9834	35.42 0.9525	39.73 0.9670	41.05 0.9679	41.05 0.9679	Python + GPU	0.18
	DRUNet	38.30 0.9608	40.33 0.9791	38.93 0.9700	41.94 0.9779	42.62 0.9819	40.35 0.9694	40.80 0.9661	40.80 0.9700	45.76 0.9895	36.61 0.9607	40.96 0.9741	42.78 0.9770	42.78 0.9770	Python + GPU	0.18

#### 4.1.2 Objective metric

Detailed average PSNR and SSIM results are given in Table 4. Overall, CBM3D is able to produce very competitive performance on almost every dataset, while DNN methods do not always demonstrate advantages over traditional denoisers, largely due to the lack of training data.

More specifically, for traditional denoisers, three different transform-domain approaches namely CBM3D, the non-local haar (NLH) transform, and color multispectral t-SVD (CMSt-SVD) show similarly good denoising performance on all datasets. Also, the effectiveness of CMSt-SVD and CBM3D1 indicates that a one-step implementation with a small number of local patches and the hard-thresholding technique is able to produce very competitive results. It is noticed that the matrix-based methods such as multi-channel weighted nuclear norm (MCWNNM) and M-SVD present almost the same denoising capability as tensor-based methods such as hyper-Laplacian regularized unidirectional low-rank tensor (LLRT) and 4DHOSVD, which indicates that the use of tensor representation may not boost structural information retrieval in realistic cases.

For DNN methods, the dual adversarial network (DANet) utilizes the PolyU dataset for training, on which it shows outstanding denoising results, but it struggles on other datasets when the training data are unavailable.

Another important and interesting observation is that DNN models trained with AWGN noise are more competitive and robust than those trained with specific real-world datasets, as can be seen from the excellent performance of FFDNet, FCCF and DRUNet on all datasets, which strongly supports the effectiveness of the Gaussian noise modeling in Eq. (1).

From the perspective of denoising speed reported in Table 4, Self2Self requires thousands of iterations to train its network with the noisy observation, resulting in the highest cost even run on advanced GPU devices. Other more efficient DNN models are able to handle color images of size  $512 \times 512 \times 3$  within 0.2 seconds. For traditional denoisers, the computational cost lies mainly in the iterative local patch search and training. For example, M-SVD spends 26 and 42 seconds on grouping and learning local SVD transform, respectively, but it is slightly faster than 4DHOSVD, because it avoids folding and unfolding operations of tensor data along different modes. Among all the traditional denoisers, the state-of-the-art CBM3D is the most efficient because it does not have to train local transforms, and its grouping step is performed only on the luminance channel.

#### 4.1.3 Visual evaluation

Visual evaluations of compared methods on CC15, PolyU and the proposed IOCI datasets are given in Fig. 8, Fig. 9 and

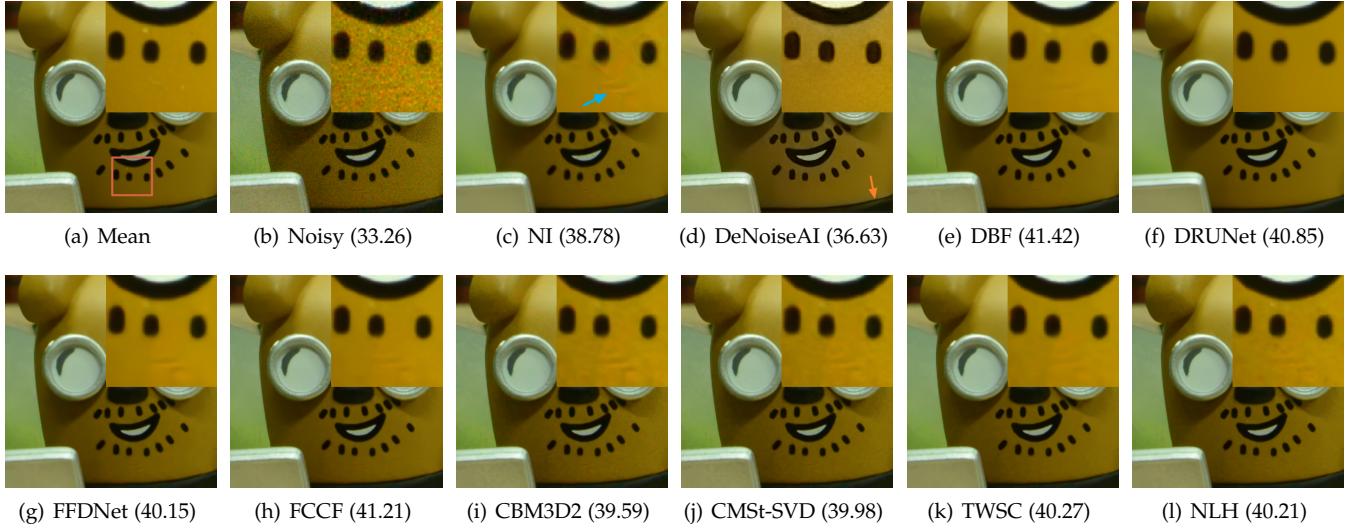


Fig. 8. Visual evaluations of compared methods (PSNR) on CC15 dataset.

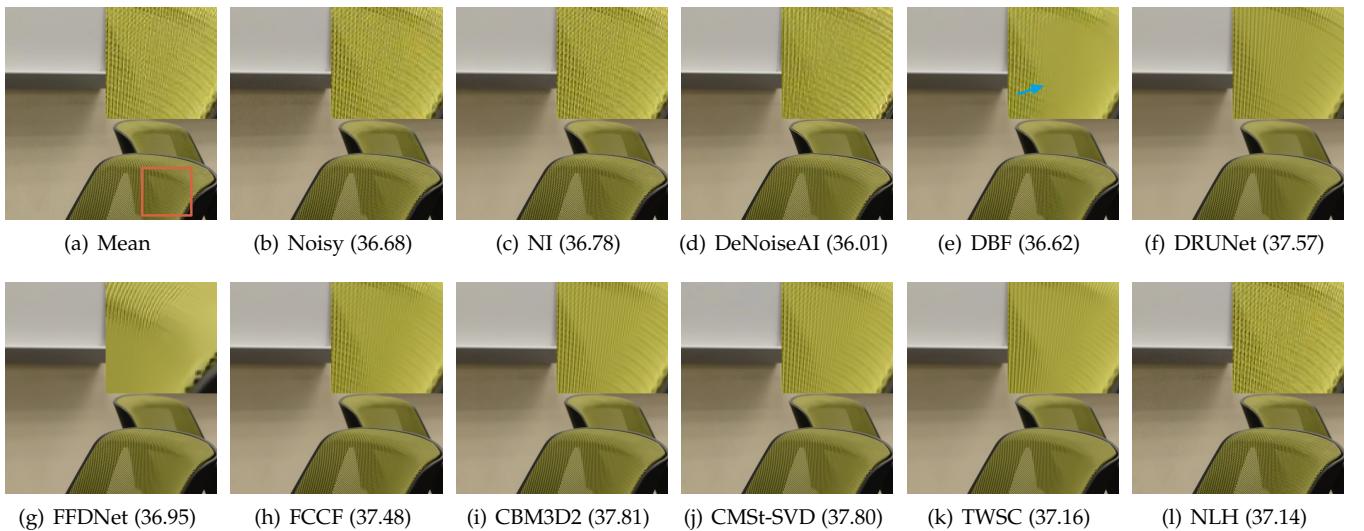


Fig. 9. Visual evaluation of compared methods (PSNR) on PolyU dataset.

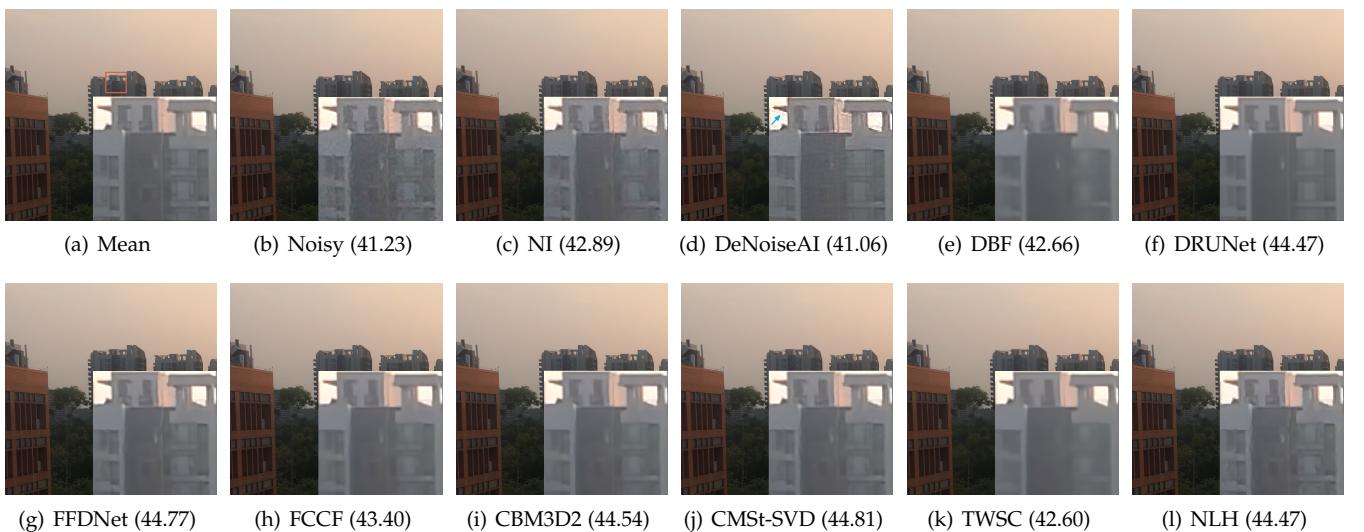


Fig. 10. Visual evaluation of compared methods (PSNR) on our IOCI's IPHONE 5S dataset.

Fig. 10, respectively. Two commercial softwares Neat Image (NI) (<https://ni.neatvideo.com/>) and Topaz DeNoiseAI (<https://topazlabs.com/>) are included in our comparison, and their optimized denoising speed enables users to fine-tune parameters for satisfactory visual effect. From Fig. 8, it can be seen that except for DRUNet and DBF, all other methods produce color artifacts, especially for patch-based approaches such as CBM3D, CMSt-SVD and NLH. Comparison in Fig. 9 indicates that an improvement of objective metric like PSNR does not always guarantee enhancement of visual effect. Except for NI, DeNoiseAI, and NLH, all other methods suffer over-smoothness to certain extents due to improper choice of parameters. From Fig. 8 and Fig. 9, it seems that DeNoiseAI improve the visual quality in terms of noise removal and detail recovery because of its unique sharpening technique. Nevertheless, it also shows some color artifacts, and an interesting example of our IOCI dataset in Fig. 10 shows that the advanced DeNoiseAI software tends to produce unwanted artifacts along the edges, which are barely noticeable in mean image, noisy observation and results of other compared methods.

## 4.2 Results for Real-World Color Video Dataset

### 4.2.1 Experimental settings

We evaluate the performance of VBM4D [17], CVMSt-SVD [19], VNLNet [33] and FastDVDNet [32] with our IOCV dataset. For VBM4D, to save some time, we only use its first stage implementation named ‘VBM4D1’. The parameters of compared methods are chosen in the same way as section 4.1.1, and similar to [32], the PSNR and SSIM of a sequence are computed as their average values of each frame.

### 4.2.2 Objective metric

Table 5 lists the average PSNR and SSIM results of compared methods. From Table 5, it can be seen that two traditional denoisers CVMSt-SVD and VBM4D show similar performance in terms of both efficiency and effectiveness. For DNN methods, VNLNet produces the best results by integrating the NLSS prior into CNN models, but its computationally expensive patch search process also results in high computational complexity. By contrast, FastDVDNet achieves almost real-time video denoising by getting rid of the time-consuming flow estimation step. Specifically, it takes FastDVDNet less than 0.1 seconds to process a single video frame, which is 8 times faster than VNLNet, and at least 13 times faster than the benchmark traditional denoisers VBM4D and CVMSt-SVD. The denoising speed of FastDVDNet is remarkable considering its competitive denoising performance in all cases. Interestingly, comparing the results of color image denoising in Table 4 and video denoising in Table 5, it is noticed that the CNN-based models present more dominating performance in the video denoising task, which may be explained by the fact that videos contain more correlated information among different frames that could be exploited by the CNN architecture.

### 4.2.3 Visual evaluation

Visual comparison is given in Fig. 11. It can be clearly observed that the noisy video is severely corrupted. The patch-based traditional denoisers CVMSt-SVD and VBM4D1 leave

TABLE 5  
PSNR and SSIM values of compared methods on real-world color video dataset. Average time per frame is also included. The best results are bolded. ‘FC’ and ‘BC’ represent front camera and back camera, respectively.

Dataset	Method				
	Noisy	CVMSt-SVD	VBM4D1	FastDVDNet	VNLNet
HUAWEI_BC	38.11	40.80	41.19	<b>41.26</b>	<b>41.35</b>
	0.9593	0.9834	0.9830	0.9857	<b>0.9868</b>
HUAWEI_FC	35.58	<b>38.71</b>	<b>38.76</b>	38.03	<b>38.79</b>
	0.9413	0.9780	0.9785	0.9776	<b>0.9809</b>
OPPO_BC	32.06	<b>33.44</b>	33.26	33.05	<b>33.56</b>
	0.9071	0.9508	0.9456	0.9476	<b>0.9544</b>
OPPO_FC	36.90	39.66	39.56	39.06	<b>40.11</b>
	0.9447	0.9791	0.9785	0.9751	<b>0.9823</b>
Implementation	/	MEX	MEX	GPU	GPU
Time/Frame (s)	/	1.86	1.29	<b>0.08</b>	0.66

behind medium-to-low-frequency noise, which would lead to video sequences with noticeable flickering, but VBM4D1 looks more visually pleasant than CVMSt-SVD because of its built-in de-flickering function. The DNN methods shows better results in this case. Specifically, VNLNet gives the best and most impressive visual effects by suppressing noise while recovering true colors and details. FastDVDNet is also able to effectively remove noise but also shows slight over-smooth effects.

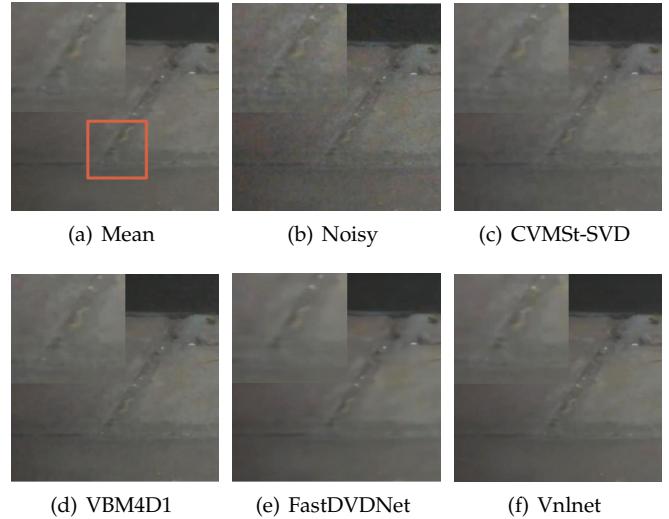


Fig. 11. Visual evaluation of color video denoising methods on our IOCV dataset.

### 4.2.4 Human ratings

Due to the limitations of hardware equipments and environment, the mean videos generated by the motorized slider with our video averaging strategy also inevitably exhibit some noise, flickering, staircase effects, motion blur and misalignment that would undermine the accuracy of objective evaluations. Also, the quality of videos may not be judged frame by frame. Therefore, we further conduct subjective evaluations by collecting human opinions [172]. More specifically, we randomly select 10 videos from our IOCV dataset, and invite 10 volunteers to rate the mean,

noisy and denoised sequences. The invited volunteers have very little background knowledge of image or video denoising, and they are not aware of how the presented video sequences are processed. For each of the 10 videos, the volunteers are asked to choose at least 2 best sequences, which then earn 1 point for the corresponding methods. The detailed human rating results are reported in Table 6. First, it is noticed that our mean-video strategy has the highest score on 9 of 10 videos, which justifies the effectiveness of the proposed IOCV dataset for objective evaluation. Furthermore, the human rating results are consistent with our evaluations in Table 5 and Fig. 11, which suggests that DNN methods output higher quality videos than the benchmark traditional denoisers on our IOCV dataset. Also, human eyes prefer VBM4D over CVMSt-SVD because CVMSt-SVD presents more temporally decorrelated low-frequency noise in flat areas, which will appear as particularly bothersome for the viewer.

TABLE 6

Human rating results of sequences generated by different methods on our IOCV dataset. The top two results are bolded.

Ratings						
# Image	Mean	Noisy	CVMSt-SVD	VBM4D1	FastDVDNet	VNLNet
1	<b>9</b>	0	1	2	<b>5</b>	3
2	<b>8</b>	1	0	4	7	6
3	<b>10</b>	0	0	8	3	5
4	<b>8</b>	0	0	4	7	<b>8</b>
5	<b>10</b>	0	0	4	4	7
6	<b>10</b>	0	0	5	6	7
7	<b>10</b>	0	0	5	<b>8</b>	5
8	<b>10</b>	2	2	8	5	6
9	<b>9</b>	0	2	4	2	7
10	4	4	4	4	5	8
Average	<b>8.80</b>	0.70	0.90	4.80	5.20	<b>6.20</b>

To conclude, the CNN-based models FastDVDNet and VNLNet show clear advantages over the representative traditional denoisers. VNLNet demonstrates state-of-the-art video denoising and enhancing performance by objective and subjective measures, and FastDVDNet performs consistently well with impressively low computational time.

### 4.3 Results for MSI Datasets

#### 4.3.1 Experimental Settings

In this section, we evaluate the performance of various MSI denoising methods on CAVE, ICVL and HHD datasets. For synthetic experiments, due to the high computational cost of some implementations, we mainly use the CAVE dataset for comparison. Similarly, considering the large size of the ICVL dataset, we select the first 20 MSI data and evaluate two representative methods MSt-SVD and QRNN3D. Apart from the classical spatial-based quality indices PSNR and SSIM, we adopt two widely used spectral-based quality indicators for MSIs, namely spectral angle mapper (SAM) [173] and dimensionless global relative error of synthesis (ERGAS) [174]. Different from PSNR and SSIM, recovered MSI data with lower SAM and ERGAS are considered of better quality.

#### 4.3.2 Synthetic MSI dataset

We assume that entries in all slices of MSI data are corrupted by zero-mean i.i.d Gaussian noise with different noise levels. Detailed denoising results on CAVE and ICVL dataset are listed in Table 7 and 8, respectively.

TABLE 8  
Average PSNR/SSIM and computational time (minutes) of MSt-SVD and QRNN3D under i.i.d Gaussian noise  $\sigma \in \{10, 20, 40\}$  on the first 20 MSI data of ICVL dataset.

Methods	$\sigma = 10$	$\sigma = 20$	$\sigma = 40$	Time (m)
MSt-SVD	46.58/0.9868	42.98/0.9746	39.32/0.9525	10.1
QRNN3D	<b>47.37/0.9908</b>	<b>45.08/0.9842</b>	<b>42.38/0.9731</b>	<b>0.16</b>

Benefiting from the advanced GPU devices, QRNN3D is 60 times faster than the benchmark traditional denoisers MSt-SVD and BM4D. However, like many DNN methods, QRNN3D is also subject to the influence of training datasets. Its models trained with the ICVL dataset can not handle noisy images from the CAVE dataset, as can be seen from the difference of Table 7 and Table 8. Similar to QRNN3D, HSI-SDeCNN also fails to produce competitive performance, because its network models are trained only with certain noise levels and bandwidths. For traditional denoisers, results in Table 7 shows the advantage of tensor-based methods by exploiting spatial and spectral correlation. Specifically, NGMeet, LLRT and LTDL consistently outperform other compared approaches at all noise levels by more than 1db. However, it takes LLRT and LTDL more than 40 minutes to process a single image, because their iterative strategy with a large number of local similar patches significantly increases the computational burden. By comparison, MSt-SVD is more efficient, since it is a one-step approach that utilizes global patch representation.

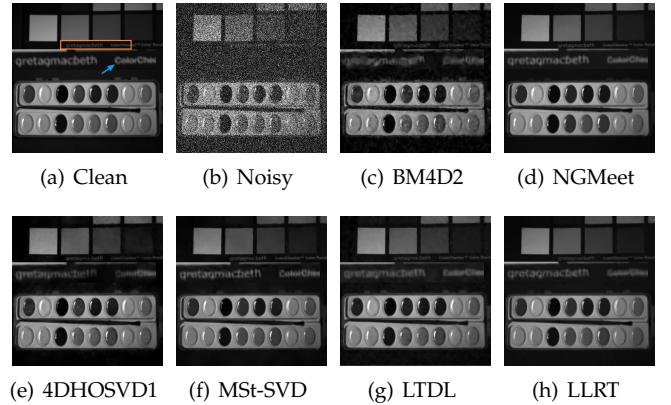


Fig. 12. Visual evaluation of MSI denoising methods on CAVE dataset under noise level  $\sigma = 100$ .

Fig. 12 compares competitive tensor-based denoisers at high noise level  $\sigma = 100$  on the CAVE dataset. In this extreme case, the pre-defined transform bases of BM4D may not take advantage of the correlation among all the spectral bands, while all other compared methods can effectively remove noise and retain details to certain extents. Interestingly, although from Table 7, MSt-SVD is outperformed by NGMeet, LLRT and LTDL, it shows the best detail recovery ability, as can be seen from the white text in Fig. 12. This

TABLE 7

Comparison of quantitative results and computational time (minutes) under i.i.d Gaussian noise  $\sigma = \{10, 30, 50, 100\}$  on CAVE dataset.

Methods	$\sigma = 10$				$\sigma = 30$				$\sigma = 50$				$\sigma = 100$				Time (m)
	PSNR	SSIM	ERGAS	SAM	PSNR	SSIM	ERGAS	SAM	PSNR	SSIM	ERGAS	SAM	PSNR	SSIM	ERGAS	SAM	
Noisy	28.13	0.4371	236.40	0.7199	18.59	0.1069	676.01	1.0085	14.15	0.0475	1126.7	1.1461	8.13	0.0136	2253.4	1.3074	/
PARAFAC	35.39	0.8759	108.78	0.2363	33.65	0.8294	125.13	0.3291	31.52	0.7393	154.19	0.4351	27.13	0.4637	250.68	0.6681	3.9
LRTA	41.36	0.9499	49.53	0.1718	36.06	0.8775	90.50	0.2446	33.52	0.8201	121.15	0.2897	30.06	0.7138	180.03	0.3649	0.5
LRMR	39.27	0.9094	64.81	0.3343	31.36	0.6451	157.65	0.6021	26.67	0.4000	264.28	0.7534	26.67	0.1850	469.26	0.9306	4.1
LLRGTV	38.96	0.9384	71.54	0.2292	33.56	0.7862	122.45	0.4756	30.33	0.6375	172.22	0.6386	25.62	0.3917	288.78	0.8593	3.8
4DHOSVD1	45.43	0.9811	30.83	0.1084	39.78	0.9336	59.12	0.2272	36.82	0.8722	83.36	0.3385	32.66	0.7307	134.34	0.5599	4.9
BM4D1	42.76	0.9716	42.19	0.1585	37.13	0.9235	82.33	0.2910	34.55	0.8803	109.43	0.3591	31.17	0.7763	158.83	0.4751	2.1
BM4D2	44.61	0.9784	33.32	0.1289	38.80	0.9283	65.23	0.2579	35.90	0.8685	91.19	0.3557	31.84	0.7197	144.91	0.5160	8.8
TDL	44.41	0.9797	34.32	0.1048	39.07	0.9493	63.18	0.1493	36.46	0.9171	85.24	0.2008	32.92	0.8284	128.15	0.3132	0.9
ITSReg	45.77	0.9802	30.53	0.1086	40.51	0.9488	53.05	0.1374	37.75	0.9271	70.16	0.1619	33.01	0.8648	120.77	0.2376	5.2
FastHyDe	42.87	0.9763	42.90	0.1398	37.22	0.9151	81.44	0.3123	34.56	0.8471	108.75	0.4395	30.62	0.6717	168.92	0.6096	0.8
HSI-SDeCNN	41.40	0.9624	52.04	0.1523	37.57	0.9247	78.28	0.1959	35.49	0.8947	98.54	0.2141	32.51	0.8309	136.16	0.2748	0.98
QRNN3D	40.36	0.9672	65.31	0.1926	38.48	0.9496	76.21	0.2320	37.02	0.9314	87.15	0.2741	30.33	0.7128	172.68	0.5161	0.01
LLRT	<b>46.60</b>	0.9868	<b>26.75</b>	<b>0.0842</b>	41.49	0.9681	48.50	<b>0.1221</b>	38.65	0.9482	67.56	<b>0.1551</b>	<b>35.39</b>	<b>0.9154</b>	<b>99.37</b>	<b>0.1962</b>	43.8
MSt-SVD	45.20	0.9814	32.05	0.1064	40.23	0.9530	56.26	0.1737	37.73	0.9285	75.03	0.2231	34.20	0.8800	115.24	0.3142	1.5
NGMeet	<b>46.58</b>	<b>0.9881</b>	27.27	0.0853	<b>41.60</b>	<b>0.9692</b>	<b>48.14</b>	0.1238	<b>38.88</b>	<b>0.9507</b>	<b>65.64</b>	0.1616	34.97	0.9072	101.98	0.2506	4.6
LTDL	45.90	0.9859	29.06	0.0925	41.20	0.9644	49.73	0.1301	38.63	0.9459	66.68	0.1650	34.87	0.9015	103.65	0.2353	60.2

observation indicates increasing the number of iterations and local similar patches may not help preserve fine details and structure of MSI data at high noise levels.

#### 4.3.3 Real-world HHD Dataset

Inevitably, images of HHD dataset are contaminated by noise to varying degrees under different illumination. Considering the large size of noisy images, we examine the effectiveness of five efficient benchmarks on handling real-world MSI data. Since there is no clean reference image, we conduct visual evaluations, and for fair comparison, the parameters of compared methods are carefully tuned. The results are visualized in Fig. 13. Without realistic training data, the CNN model of QRNN3D produces obviously blurry image. Interestingly, different from our observations of synthetic AWGN comparison in Fig. 12, no compared methods could consistently outperform BM4D in this realistic case. As can be seen in Fig. 13, all state-of-the-art algorithms struggle to adapt to different local image contents to balance smoothness and sharp details. For patch-based methods, applying a same set of parameters or transforms to multiple spectral bands may result in similar denoise patterns across the whole image. Therefore, apart from the high computational burden, another challenge of filtering MSI data lies in dealing with the spatial and spectral variation of noise levels and distributions.

#### 4.4 Results for MRI datasets

##### 4.4.1 Experimental Settings

Different from the AWGN noise modeling of MSI, MRI data are often corrupted by Rician noise [40]. Specifically, Let  $\mathcal{X}$  be the original noise-free signal, the noise Rician MRI data  $\mathcal{Y}$  can be obtained via

$$\mathcal{Y} = \sqrt{(\mathcal{X} + \sigma n_1)^2 + (\sigma n_2)^2} \quad (9)$$

where  $n_1, n_2 \sim N(0, 1)$  are i.i.d. random vectors following the standard normal distribution,  $\sigma$  is the standard deviation in both real and imaginary channels of the noisy 3D MR

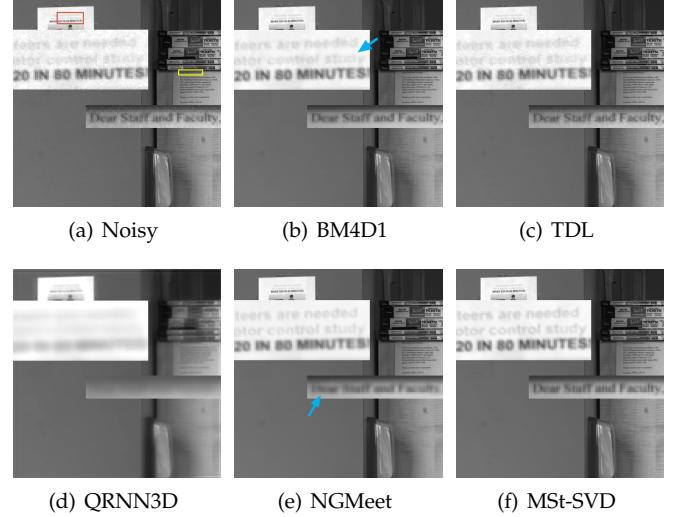


Fig. 13. Visual evaluation of compared MSI denoising methods on the real-world HHD dataset.

images. To handle Rician noise, the technique of forward and inverse variance stabilizing transform (VST) [175] is often adopted by Gaussian denoisers via

$$\hat{\mathcal{X}} = VST^{-1}(denoise(VST(\mathcal{Y}, \sigma), \sigma_{VST}), \sigma) \quad (10)$$

where  $VST^{-1}$  denotes the inverse of  $VST$ ,  $\sigma_{VST}$  is the stabilized standard deviation after  $VST$  transform, and  $\sigma$  is the standard deviation of the noise in Eq. (9). According to Eq. (10), the noisy Rician data  $\mathcal{Y}$  is first stabilized by the  $VST$  and then filtered by certain Gaussian denoisers using a constant noise level  $\sigma_{VST}$ , and then the final estimate is obtained by applying the inverse  $VST$  to the output of the denoising result [13].

In our experiments, we use PSNR and SSIM as the objective metrics, and similar to [13], [70], the PSNR value is only computed on the foreground defined by  $\mathcal{X}_f = \{x \in \mathcal{X} : x > 10 \cdot D/255\}$ , where  $D$  is the peak of clean data  $\mathcal{X}$ .

#### 4.4.2 Synthetic Brainweb dataset

The volume data are added with varying levels of stationary Rician noise from 1% to 19% of the maximum intensity with an increase of 2 %. In real world cases, the noise level is usually much lower than 19%, but we are also interested in the denoising capability of compared methods under extreme conditions. Table 9 lists detailed quantitative results, and Fig. 14 compares the denoising performance at high noise levels when  $\sigma \geq 11\%$ . At lower noise levels, PRI-NLPCA is able to take advantage of the high-quality initial estimate of NLPCA, and achieves the best performance when  $\sigma \leq 9\%$ . As noise level increases, tensor-based methods demonstrate advantages of extracting latent features, and the iterative low-rank HOSVD (ILR-HOSVD) produces the best results when  $11\% \leq \sigma \leq 15\%$ . However, the iterative learning strategy is subject to the presence of extremely high noise, while BM4D benefits from its pre-defined transforms and outperforms other methods when  $\sigma \geq 17\%$ .

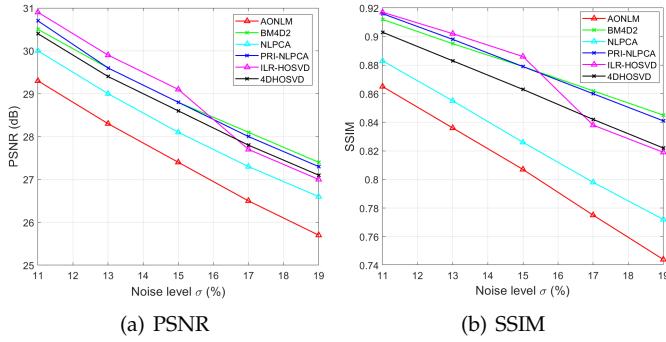


Fig. 14. Average PSNR and SSIM values of representative compared methods with T1w, T2w and PDw data at high noise levels ( $\sigma \geq 11\%$ ).

Visual evaluation on T1w data when  $\sigma = 19\%$  is illustrated in Fig. 15, and it shows that BM4D2 is successful in both noise suppression and detail preservation. Compared to NLPCA and ILR-HOSVD, MSt-SVD is less affected by severe noise, as it may benefit from global patch representation with fast Fourier transform (FFT). The state-of-the-art NLM methods, namely PRINLM and PRI-NLPCA exhibit pleasant visual effects in homogeneous areas at the cost of slight over-smoothness along the edges. Furthermore, PRI-NLPCA shows some artifacts on the background due to the influence of noise on learning local PCA transform.

#### 4.4.3 Real-world MRI dataset

To evaluate the performance of compared methods on real-world 3D MR data, we carry out experiments on T1w MR images of the OASIS dataset [165]. The Rician noise levels of two selected T1w data, namely OAS1\_0112 and OAS1\_0092 are estimated to be 3% and 4.5% of the maximum intensity, respectively [78]. The filtered images of different methods are compared in Fig. 16 and Fig. 17, respectively. Based on our synthetic experiments in Table 9, we learn that in most cases all methods could achieve competitive denoising performance at low noise levels, as can also be seen from the excellent visual effects on realistic OAS1\_0112 data in Fig. 16. Thus, from the perspective of real-world MRI denoising,

BM4D1 and PRINLM should be preferred due to their low computational cost reported in Table 9. Despite the similar denoising results of compared methods, we also observe that on certain slices of the 3D MR data, NLPCA produces the best results in terms of fine detail preservation, while other methods exhibit over-smoothness to varying degrees, as is illustrated in Fig. 17. This interesting observation is another vivid example to show that the use of tensor representation does not always help preserve more structural information of the underlying clean images.

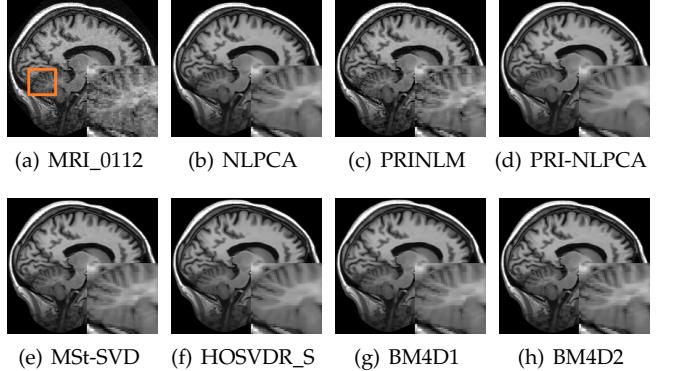


Fig. 16. Visual evaluation of compared methods on Real OAS1\_0112 T1w data with estimated noise level  $\sigma = 3\%$ .

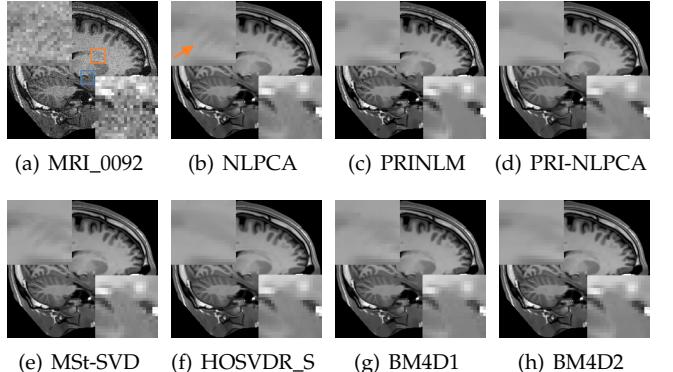


Fig. 17. Visual evaluation of compared methods on Real OAS1\_0092 T1w data with estimated noise level  $\sigma = 4.5\%$ .

## 4.5 Discussion

Our discussion in this section will mainly focus on color image denoising due to the abundance of advanced compared methods and real-world datasets.

### 4.5.1 The robustness of CBM3D

Recently, there is some criticism [127] of CBM3D that to achieve competitive denoising performance, it may need to traverse different Gaussian noise levels  $\sigma$ , which could be potentially time-consuming and impractical without ground-truth. Such criticism may be unfair for CBM3D for two reasons. First, compared to nearly all methods listed in Table 4, CBM3D does not involve any training process which could be influenced by the noise distributions of specific datasets. Second, in our experiments, we discover that the pre-defined transforms of CBM3D are not very sensitive to the choice of the input noise level parameter  $\sigma$ .

TABLE 9

Average PSNR/SSIM and computational time (s) of different methods with T1w, T2w, and PDw data corrupted by Rician noise. The standard deviations  $\sigma$  of the noise is expressed as percentage relative to the maximum intensity value of the noise-free data [13].

Noise	Noisy	ONLM	AONLM	ODCT	PRINLM	BM4D1	BM4D2	NLPCA	PRI-NLPCA	ILR-HOSVD	4DHOSVD	MSt-SVD
1	40.0/0.971	41.7/0.991	41.9/0.991	43.3/0.992	43.3/0.993	43.2/0.992	43.4/0.992	44.1/0.994	44.3/0.994	<b>44.4/0.994</b>	43.9/0.993	44.0/0.993
3	30.5/0.822	36.4/0.969	36.5/0.971	36.4/0.970	36.9/0.975	36.6/0.972	37.2/0.975	37.7/0.977	<b>38.1/0.981</b>	38.0/0.978	37.5/0.976	37.5/0.975
5	26.1/0.676	33.3/0.940	33.8/0.947	33.6/0.949	34.0/0.955	33.7/0.950	34.6/0.959	33.8/0.940	<b>35.3/0.967</b>	35.2/0.963	34.6/0.959	34.6/0.958
7	23.2/0.564	31.2/0.911	31.9/0.921	31.7/0.927	32.0/0.934	31.8/0.926	32.8/0.943	31.9/0.912	<b>33.4/0.951</b>	33.4/0.948	32.9/0.941	32.8/0.940
9	21.0/0.479	29.4/0.880	30.5/0.894	30.2/0.905	30.5/0.911	30.3/0.899	31.5/0.928	31.2/0.908	<b>31.9/0.934</b>	<b>32.0/0.933</b>	31.5/0.922	31.3/0.924
11	19.3/0.414	27.8/0.847	29.3/0.865	29.1/0.883	29.3/0.888	29.2/0.873	30.5/0.912	30.0/0.883	30.7/0.916	<b>30.9/0.917</b>	30.4/0.903	30.2/0.906
13	17.8/0.361	26.3/0.811	28.3/0.836	28.1/0.861	28.2/0.864	28.1/0.844	29.6/0.895	29.0/0.855	29.6/0.898	<b>29.9/0.902</b>	29.4/0.883	29.3/0.888
15	16.6/0.319	24.9/0.771	27.4/0.807	27.3/0.840	27.4/0.841	27.3/0.814	28.8/0.879	28.1/0.826	28.8/0.879	<b>29.1/0.886</b>	28.6/0.863	28.5/0.871
17	15.5/0.284	23.4/0.729	26.5/0.775	26.5/0.820	26.6/0.817	26.5/0.785	<b>28.1/0.862</b>	27.3/0.798	28.0/0.860	27.7/0.838	27.8/0.842	27.7/0.853
19	14.6/0.254	22.0/0.684	25.7/0.744	25.8/0.799	25.9/0.797	25.7/0.755	<b>27.4/0.845</b>	26.6/0.772	27.3/0.841	27.0/0.819	27.1/0.822	27.1/0.835
Implementation	/	Mex	Mex	Mex	Mex	Mex	MATLAB	MEX	MATLAB	MATLAB	MEX	
Time (s)	/	139.9	868.1	<b>10.5</b>	18.6	69.9	211.3	638.8	868.9	1896.3	706.9	99.6

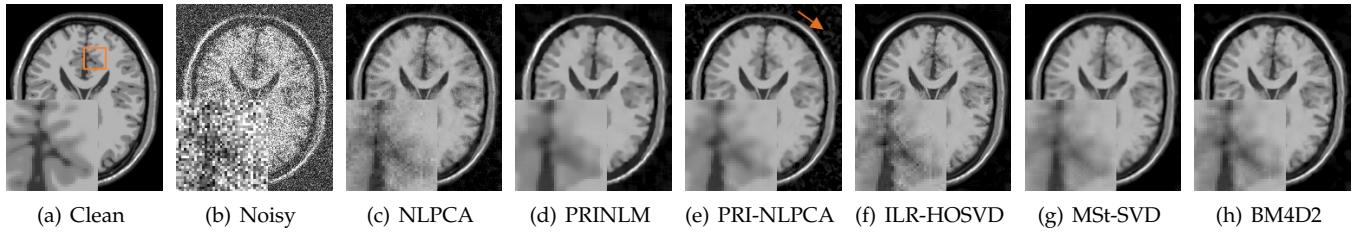


Fig. 15. Visual evaluation of compared methods on synthetic Brainweb T1w data with noise level  $\sigma = 19\%$ .

within a reasonable range. Fig. 18 compares the PSNR and SSIM values of CBM3D1 with  $\sigma \in [10, 30]$  on five datasets. It is noticed that choosing a noise level  $\sigma$  between 10 and 20 could yield similarly good performance for CBM3D, making it an exceptionally robust method. Furthermore, from the perspective of blind denoising [68], [121], [176], [177], CBM3D could be viewed as a competitive blind denoiser with input noise level fixed to be  $\sigma = 15$ .

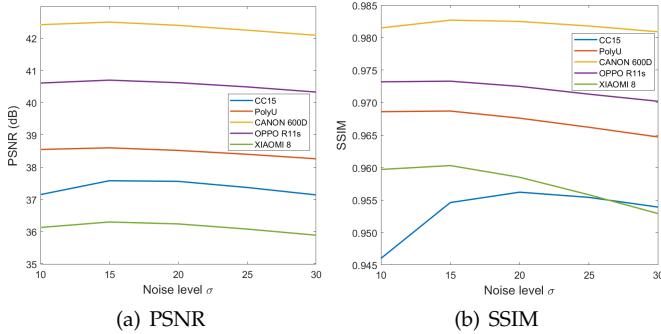


Fig. 18. PSNR and SSIM change of CBM3D1 with noise level  $\sigma \in [10, 30]$  on five real-world color image datasets.

#### 4.5.2 Understanding the denoising performance

The theoretical bound of compared image denoising methods is hard to obtain, but it is interesting to investigate the denoising capability of compared methods under the challenging practical cases when prior information, such as training datasets and camera settings are unavailable. We use our IOCI's FUJIFILM X100T camera dataset, and for each scene, we generate three new mean images by averaging 3, 5 and 10 noisy images, and they are named

'Mean\_3', 'Mean\_5' and 'Mean\_10', respectively. They could be regarded as images captured via different continuous shooting modes with 'high', 'medium' and 'low' noise levels. Fig. 19 illustrates the average PSNR differences of six implementations compared with 'Mean\_3' on FUJI dataset. Interestingly, Fig. 19 shows that state-of-the-art denoising methods only produce marginal improvements compared with Mean\_3, which indicates that their denoising performance is similar to obtaining the mean image by averaging 3 consecutive noisy observations.

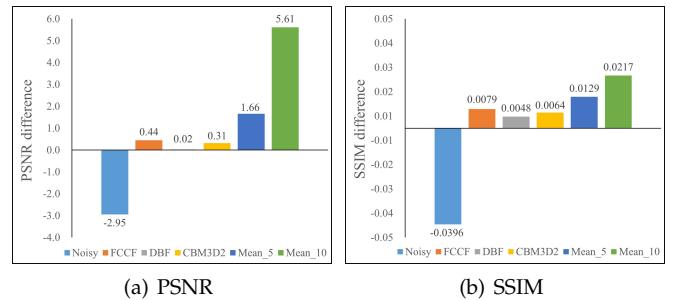


Fig. 19. PSNR and SSIM difference of six implementations compared with 'Mean\_3' on FUJI dataset.

#### 4.5.3 Denoising with resizing

To boost the denoising performance of traditional patch-based denoisers, many methods [11], [63], [64], [74], [77], [78], [84], [86] propose to iteratively combine the filtered image with the corresponding noisy observation in different ways [41], [178], [179], [180]. A simple approach is to add back the noisy observation  $\mathcal{Y}$  to the denoised image  $\hat{\mathcal{X}}$  via

$$\hat{\mathcal{Y}} = \lambda \mathcal{Y} + (1 - \lambda) \hat{\mathcal{X}} \quad (11)$$

where  $\lambda \in [0, 1]$  is a relaxation parameter reflecting the importance of  $\mathcal{Y}$ . The new noisy image  $\hat{\mathcal{Y}}$  is then fed into the denoiser for final estimate. The iterative denoising strategies may significantly increase the computational burden, and the results reported in Table 7 and Table 9 show that a simple one-stage implementation such as MSt-SVD and BMXD1 could produce competitive performance at moderate noise levels. It is therefore interesting to investigate how to efficiently handle severe noise with traditional denoisers. Recently, Zontak et al. [181] show that in the down-sampled noisy image, patches tend to be noiseless and share similar patterns with underlying clean patches. Therefore, instead of directly filtering the large-size noisy observation, an alternative is to first handle downsized image, and then upscale the denoised image back to its original size with some effective image super-resolution algorithms [182], [183]. This idea is similar to the encoder-decoder scheme of DNN methods. For simplicity we use MATLAB's built-in 'imresize' function and applies it to the efficient CMSt-SVD. This special implementation is termed as 'CMSt-SVD\_R'. The visual effects of CMSt-SVD with and without the image resizing strategy are compared in Fig. 20. As can be seen, the resizing strategy could effectively reduce color artifacts at the cost of slight over-smooth effects, thus it may be more effective and suitable for very noisy images with fewer fine details and textures.

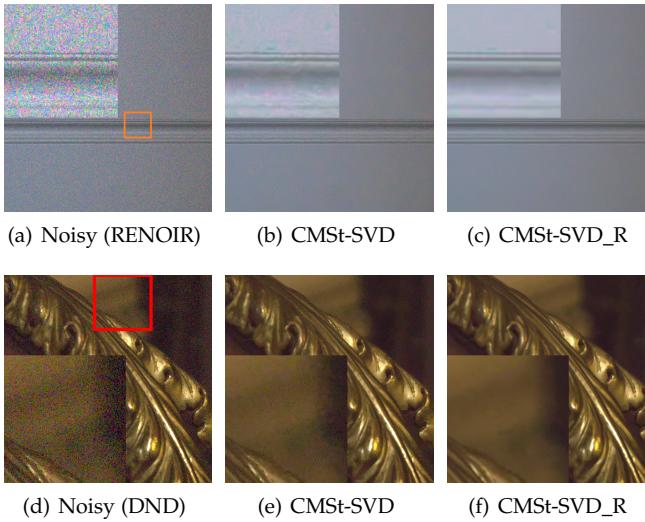


Fig. 20. Visual evaluation of CMSt-SVD with and without the image resizing strategy on RENOIR and DND datasets.

## 5 CONCLUSION

Recently, multi-dimensional image denoising has become an attractive research topic with potentially important applications. The great success of the BM3D-based methods contributes significantly to the emergence of numerous related denoising methods, varying from traditional Gaussian denoisers to advanced DNN methods. In this paper, new benchmark datasets for color image and video are introduced, and comprehensive comparison of representative denoising methods are conducted with both synthetic and real-world experimental settings. Throughout extensive objective and subjective evaluations, we show: (i) the overall powerful denoising ability of the BM3D family, (ii) the use of

tensor representation does not guarantee better results compared to matrix implementations, and (iii) the outstanding generalizability and denoising performance of some DNN methods such as FCCF, DRUNet, FastDVDNet and VNLNet to real-world color image/video datasets.

Despite the achievements in recent years, the state-of-the-art approaches still suffer from limitations that restrict their application to key real-world scenarios. For example, from Fig. 9 and Fig. 10, it is noticed that many image denoising methods show only limited enhancement and may even degrade the image quality with over-smoothness and annoying artifacts. As image size grows rapidly with the development of new imaging technologies, a challenging problem is to decide if it is worth applying filtering to an image at hand [184], and if denoising is desired, which of the filters listed in Table 1 and Table 2 should be used. Recently, a new trend focuses on how to jointly handle image denoising and other computer vision tasks such as dehazing [185], demosaicing [186] and classification [187]. It is interesting to collect realistic benchmark datasets to further explore their mutual influence.

## APPENDIX A THE MODIFIED SVD METHOD

In this section, we introduce the motivation and details of the modified SVD (M-SVD) method briefed in Algorithm 1. For simplicity, our analysis is based on the color image denoising task. Given the local group  $\mathcal{G} \in \mathbb{R}^{ps \times ps \times 3 \times K}$  of  $K$  patches, the original SVD algorithm operates directly on the matrix representation  $\mathbf{G}_{(4)}$  of  $\mathcal{G}$  via

$$\mathbf{G}_{(4)} = \mathbf{U}\mathbf{S}\mathbf{V}^T \quad (12)$$

where  $\mathbf{S}$  is singular value matrix,  $\mathbf{U}$  and  $\mathbf{V}$  are group and patch level transform matrices, respectively. A more effective way to obtain  $\mathbf{U}$  is to perform SVD on the opponent color space [11], [19] represented by  $\mathbf{G}_{opp_{(3)}}$ , with  $\mathcal{G}_{opp} = \mathcal{G}(:, :, 1, :) + \mathcal{G}(:, :, 2, :) + \mathcal{G}(:, :, 3, :)$ . Then the core matrix  $\mathbf{C}$  of  $\mathbf{G}_{(4)}$  can be calculated by

$$\mathbf{C} = \mathbf{U}^T \mathbf{G}_{(4)} \mathbf{V} \quad (13)$$

After the truncation of  $\mathbf{C}$  and the inverse transform of (13), the estimated clean group  $\hat{\mathbf{G}}_c = \mathbf{U}\mathbf{C}_{trun}\mathbf{V}^T$  is obtained.

---

### Algorithm 1 M-SVD

**Input:** Noisy image  $\mathcal{A}$ , number of similar patches  $K$ .

**Output:** Estimated clean image  $\hat{\mathcal{A}}_c$ .

**Step 1** (Grouping): For every reference patch of  $\mathcal{A}$ , stack  $K$  similar patches in a group  $\mathcal{G} \in \mathbb{R}^{ps \times ps \times 3 \times K}$ .

**Step 2** (Collaborative filtering):

(1) Obtain the group and patch level transform matrices  $\mathbf{U}$  and  $\mathbf{V}$  by performing SVD on  $\mathbf{G}_{opp_{(3)}}$  and  $\mathbf{G}_{(4)}^T$ , respectively.

(2) Apply the hard-threshold technique to  $\mathbf{C} = \mathbf{U}^T \mathbf{G}_{(4)} \mathbf{V}$  via  $\mathbf{C}_{trun} = \text{hard-threshold}(\mathbf{C})$ .

(3) Take the inverse transform to obtain estimated clean group via  $\hat{\mathbf{G}}_c = \mathbf{U}\mathbf{C}_{trun}\mathbf{V}^T$ .

**Step 4** (Aggregation): Averagely write back all image patches in  $\hat{\mathbf{G}}_c$  to their original locations according to Eq. (7).

---

## APPENDIX B

### A CLOSER LOOK AT THE EFFECTIVENESS OF HOSVD FOR DENOISING

Our real-world experiments of color image and MRI denoising in section 4.1 and 4.4 show that the SVD-based methods could produce very competitive performance compared with HOSVD-based approaches, which are also prone to over-smooth textures. With such observation, we further investigate and discuss the effectiveness of HOSVD for denoising. Our analysis in this section is developed mainly based on the results of [34], [98], [104], [188], [189], [190], with more details available in the supplementary file.

#### B.1 The relationship between HOSVD and SVD

The relationship between HOSVD and SVD is established mainly based on the following Theorems.

**Theorem 1.** Given an  $N$ -th order tensor  $\mathcal{G} \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_N}$ , its full HOSVD is

$$\mathcal{G} = \mathcal{C} \times_1 \mathbf{U}_1 \times_2 \mathbf{U}_2 \times \dots \times_N \mathbf{U}_N \quad (14)$$

where  $\mathcal{C} \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_N}$  represents core tensor. If  $\mathbf{S}_i$  is the singular value matrix of  $\mathbf{G}_{(i)}$ , then

$$\mathbf{C}_{(i)} \mathbf{C}_{(i)}^T = \mathbf{S}_i \mathbf{S}_i^T \quad (15)$$

**Theorem 2.** Given an  $N$ -th order tensor  $\mathcal{G} \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_N}$  and its core tensor  $\mathcal{C}$  of the same size, with  $\mathbf{G}_{(i)} = \mathbf{U}_i \mathbf{S}_i \mathbf{V}_i^T$  via SVD, then the  $j$ -th column of  $\mathbf{V}_i$  can be represented as

$$\mathbf{V}_i(:, j) = \frac{\hat{\mathbf{U}}_i \mathbf{C}_{(i)}(j, :)^T}{\|\mathbf{C}_{(i)}(j, :)^T\|} \quad (16)$$

where  $\hat{\mathbf{U}}_i = \mathbf{U}_N \otimes \dots \otimes \mathbf{U}_{i+1} \otimes \mathbf{U}_{i-1} \otimes \dots \otimes \mathbf{U}_1$ .

For simplicity, consider the case of third-order tensor group  $\mathcal{G} \in \mathbb{R}^{I_1 \times I_2 \times K}$  with  $K$  patches of size  $I_1 \times I_2$ , its full HOSVD is given by

$$\mathcal{G} = \mathcal{C} \times_1 \mathbf{U}_1 \times_2 \mathbf{U}_2 \times_3 \mathbf{U}_3 \quad (17)$$

The full SVD of the matrix representation  $\mathbf{G}_{(3)}$  is

$$\mathbf{G}_{(3)} = \mathbf{U}_3 \mathbf{S}_3 \mathbf{V}_3^T \quad (18)$$

Theorem 1 indicates that truncating the columns of the group-level transform  $\mathbf{U}_3$  is equivalent to shrinking the rows of  $\mathbf{C}_{(3)}$  and also the singular value matrix  $\mathbf{S}_3$  of  $\mathbf{G}_{(3)}$ . Theorem 2 demonstrates how the patch-level transform  $\mathbf{V}_3$  of  $\mathbf{G}_{(3)}$  can be derived from the first and second mode projection matrices  $\mathbf{U}_1$  and  $\mathbf{U}_2$  of  $\mathcal{G}$ .

#### B.2 The risk of truncated HOSVD for denoising

HOSVD-based methods remove noise by truncating the columns of projection matrices via low-rank approximation, or by shrinking the core tensor via the hard-thresholding technique. To demonstrate the over-smooth effects of truncated HOSVD (T-HOSVD), we use a noise-free group  $\mathcal{G} \in \mathbb{R}^{3 \times 3 \times 3}$  consisting of three same patches with  $\mathcal{G}(:, :, 1) = \mathcal{G}(:, :, 2) = \mathcal{G}(:, :, 3)$ , and the third-mode unfolding of  $\mathcal{G}$  is

$$\mathbf{G}_{(3)} = \begin{pmatrix} 3 & 1 & 5 & 6 & 4 & 8 & 9 & 2 & 6 \\ 3 & 1 & 5 & 6 & 4 & 8 & 9 & 2 & 6 \\ 3 & 1 & 5 & 6 & 4 & 8 & 9 & 2 & 6 \end{pmatrix} \quad (19)$$

where every row of  $\mathbf{G}_{(3)}$  corresponds to a vectorized image patch. According to Eq. (17), the core tensor  $\mathcal{C}$  is

$$\mathbf{C}_{(3)} = \begin{pmatrix} -27.98 & 0 & 0 & 0 & -5.38 & 0 & 0 & 0 & 2.07 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} \quad (20)$$

Assume the pre-defined multi-rank  $r_1 = r_2 = 2$  is applied to truncate  $\mathbf{U}_1$  and  $\mathbf{U}_2$  with  $\mathbf{U} = \mathbf{U}(:, 1 : 2)$ , then after the inverse transform of Eq. (17), the corresponding denoised group  $\hat{\mathcal{G}}_c$  is

$$\begin{pmatrix} 3.09 & 1.92 & 4.54 & 5.95 & 3.46 & 8.26 & 9.00 & 2.03 & 5.98 \\ 3.09 & 1.92 & 4.54 & 5.95 & 3.46 & 8.26 & 9.00 & 2.03 & 5.98 \\ 3.09 & 1.92 & 4.54 & 5.95 & 3.46 & 8.26 & 9.00 & 2.03 & 5.98 \end{pmatrix} \quad (21)$$

Comparing Eq. (19) and Eq. (21), the loss of information is  $\|\mathbf{G}_{(3)} - \hat{\mathbf{G}}_{c(3)}\|_F^2 = 4.29$ . Also, from (20) it is noticed that the hard-thresholding technique in Eq. (5) will yield the same loss if the threshold parameter  $\tau$  is set to  $2.07 < \tau < 5.38$ . Fig. 21 illustrates the over-smooth effects of T-HOSVD when applied to a noise-free image with all the patches in a group forced to be the same.

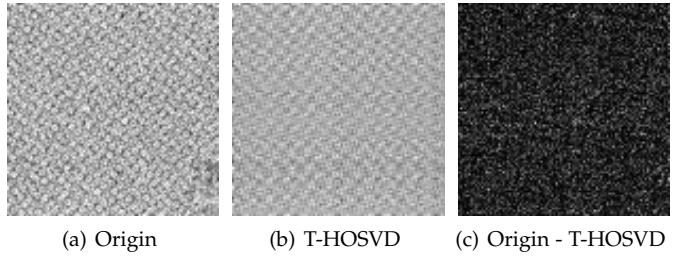


Fig. 21. Illustration of over-smooth effects produced by truncated HOSVD (T-HOSVD) when all the patches in a group are the same.

Our analysis in this section is based on an extreme and ideal case that all patches in a noise-free group are the same. In real-world applications, the difference among nonlocal similar patches is not negligible, and the use of tensor decomposition techniques may help filter out more redundant information and extract latent features, as is observed in the MRI denoising experiment illustrated in Fig. 15. Also, it is noticed from Fig. 21 that certain structural shapes and textures are preserved. However, despite recent development of the tensor theory [191], the choice of the best multi-rank and threshold parameter  $\tau$  for the tensor truncation strategy still remains a challenge.

## ACKNOWLEDGMENT

The authors would like to thank those researchers who make their code, software packages and datasets publicly available. We also appreciate the generosity of our friends and volunteers for sharing their camera devices and participating in the video quality assessment task.

## REFERENCES

- [1] T. Tirer and R. Giryes, "Back-projection based fidelity term for ill-posed linear inverse problems," *IEEE Trans. Image Process.*, vol. 29, pp. 6164–6179, 2020.
- [2] M. Isard and A. Blake, "Condensation conditional density propagation for visual tracking," *Int. J. Comput. Vis.*, vol. 29, no. 1, pp. 5–28, 1998.

- [3] N. R. Pal and S. K. Pal, "A review on image segmentation techniques," *Pattern Recognit.*, vol. 26, no. 9, pp. 1277–1294, 1993.
- [4] M. K. Raczkowska, P. Koziol, S. Urbaniak-Wasik, C. Palusziewicz, W. M. Kwiatek, and T. P. Wrobel, "Influence of denoising on classification results in the context of hyperspectral data: High definition ft-ir imaging," *Anal. Chim. Acta*, vol. 1085, pp. 39–47, 2019.
- [5] N. Wiener, *Extrapolation, interpolation, and smoothing of stationary time series: with engineering applications*. MIT press, 1950.
- [6] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising by sparse 3-d transform-domain collaborative filtering," *IEEE Trans. Image Process.*, vol. 16, no. 8, pp. 2080–2095, 2007.
- [7] A. Buades, B. Coll, and J.-M. Morel, "A review of image denoising algorithms, with a new one," *Multiscale Model. Simul.*, vol. 4, no. 2, pp. 490–530, 2005.
- [8] L. P. Yaroslavsky, K. O. Egiazarian, and J. T. Astola, "Transform domain image restoration methods: review, comparison, and interpretation," in *Proc. Nonlinear Image Process. Pattern Anal. XII*, vol. 4304, 2001, pp. 155–169.
- [9] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Bm3d image denoising with shape-adaptive principal component analysis," in *Proc. Workshop on SPARS*, 2009, pp. 1–6.
- [10] V. Katkovnik, A. Foi, K. Egiazarian, and J. Astola, "From local kernel to nonlocal multiple-model image denoising," *Int. J. Comput. Vis.*, vol. 86, no. 1, pp. 1–32, 2010.
- [11] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Color image denoising via sparse 3d collaborative filtering with grouping constraint in luminance-chrominance space," in *Proc. IEEE Int. Conf. Image Process.*, 2007, pp. 313–316.
- [12] A. Danielyan, A. Foi, V. Katkovnik, and K. Egiazarian, "Denoising of multispectral images via nonlocal groupwise spectrum-pca," in *Proc. Conf. Colour Graph., Imag. Vis.*, vol. 2010, no. 1, 2010, pp. 261–266.
- [13] M. Maggioni, V. Katkovnik, K. Egiazarian, and A. Foi, "Nonlocal transform-domain filter for volumetric data denoising and reconstruction," *IEEE Trans. Image Process.*, vol. 22, no. 1, pp. 119–133, 2012.
- [14] L. Shao, R. Yan, X. Li, and Y. Liu, "From heuristic optimization to dictionary learning: A review and comprehensive comparison of image denoising algorithms," *IEEE Trans. Cybern.*, vol. 44, no. 7, pp. 1001–1013, 2013.
- [15] C. Tian, L. Fei, W. Zheng, Y. Xu, W. Zuo, and C.-W. Lin, "Deep learning on image denoising: An overview," *Neural Networks*, vol. 131, pp. 251–275, 2020.
- [16] D. Kostadin, F. Alessandro, and E. Karen, "Video denoising by sparse 3d transform-domain collaborative filtering," in *Proc. 15th Eur. Signal Process. Conf.*, 2007, pp. 145–149.
- [17] M. Maggioni, G. Boracchi, A. Foi, and K. Egiazarian, "Video denoising, deblocking, and enhancement through separable 4-d nonlocal spatiotemporal transforms," *IEEE Trans. Image Process.*, vol. 21, no. 9, pp. 3952–3966, 2012.
- [18] T. Plotz and S. Roth, "Benchmarking denoising algorithms with real photographs," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 1586–1595.
- [19] Z. Kong and X. Yang, "Color image and multispectral image denoising using block diagonal representation," *IEEE Trans. Image Process.*, vol. 28, no. 9, pp. 4247–4259, 2019.
- [20] X. Liu, M. Tanaka, and M. Okutomi, "Single-image noise level estimation for blind denoising," *IEEE Trans. Image Process.*, vol. 22, no. 12, pp. 5226–5237, 2013.
- [21] G. Chen, F. Zhu, and P. Ann Heng, "An efficient statistical method for image noise level estimation," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2015, pp. 477–485.
- [22] B. Wen, "Reproducible denoising methods," [Online]. Available: <https://github.com/wenbihan/reproducible-image-denoising-state-of-the-art>.
- [23] P. Milanfar, "A tour of modern image filtering: New insights and methods, both practical and theoretical," *IEEE Sig. Process. Mag.*, vol. 30, no. 1, pp. 106–128, 2012.
- [24] D. Thanh, P. Surya *et al.*, "A review on ct and x-ray images denoising methods," *Informatica*, vol. 43, no. 2, 2019.
- [25] J. Mohan, V. Krishnaveni, and Y. Guo, "A survey on the magnetic resonance image denoising methods," *Biomed. Sign. Process. Control*, vol. 9, pp. 56–69, 2014.
- [26] J. Schmidhuber, "Deep learning in neural networks: An overview," *Neural networks*, vol. 61, pp. 85–117, 2015.
- [27] Z. Kong and X. Yang, "A brief review of real-world color image denoising," *arXiv preprint arXiv:1809.03298*, 2018.
- [28] H. Yue, J. Liu, J. Yang, T. Q. Nguyen, and F. Wu, "High iso jpeg image denoising by deep fusion of collaborative and convolutional filtering," *IEEE Trans. Image Process.*, vol. 28, no. 9, pp. 4339–4353, 2019.
- [29] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Medical Image Comput. Comput.-Assisted Intervention*, 2015, pp. 234–241.
- [30] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 770–778.
- [31] K. Zhang, Y. Li, W. Zuo, L. Zhang, L. Van Gool, and R. Timofte, "Plug-and-play image restoration with deep denoiser prior," *arXiv preprint arXiv:2008.13751*, 2020.
- [32] M. Tassano, J. Delon, and T. Veit, "Fastdvdnet: Towards real-time deep video denoising without flow estimation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, June 2020.
- [33] A. Davy, T. Ehret, J.-M. Morel, P. Arias, and G. Facciolo, "Non-local video denoising by cnn," *arXiv preprint arXiv:1811.12758*, 2018.
- [34] T. G. Kolda and B. W. Bader, "Tensor decompositions and applications," *SIAM Rev.*, vol. 51, no. 3, pp. 455–500, 2009.
- [35] R. Ramanath, W. E. Snyder, Y. Yoo, and M. S. Drew, "Color image processing pipeline," *IEEE Sig. Process. Mag.*, vol. 22, no. 1, pp. 34–43, 2005.
- [36] M. Yan, "Restoration of images corrupted by impulse noise and mixed gaussian impulse noise using blind inpainting," *SIAM J. Imag. Sci.*, vol. 6, no. 3, pp. 1227–1245, 2013.
- [37] 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, 2017.
- [38] Z. Zhang, G. Ely, S. Aeron, N. Hao, and M. Kilmer, "Novel methods for multilinear data completion and de-noising based on tensor-svd," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2014, pp. 3842–3849.
- [39] Y. Chen, X. Cao, Q. Zhao, D. Meng, and Z. Xu, "Denoising hyperspectral image with non-iid noise structure," *IEEE Trans. Cybern.*, vol. 48, no. 3, pp. 1054–1066, 2017.
- [40] S. P. Awate and R. T. Whitaker, "Feature-preserving mri denoising: a nonparametric empirical bayes approach," *IEEE Trans. Med. Imag.*, vol. 26, no. 9, pp. 1242–1255, 2007.
- [41] Y. Romano, M. Elad, and P. Milanfar, "The little engine that could: Regularization by denoising (red)," *SIAM J. Imag. Sci.*, vol. 10, no. 4, pp. 1804–1844, 2017.
- [42] J. Xu, L. Zhang, W. Zuo, D. Zhang, and X. Feng, "Patch group based nonlocal self-similarity prior learning for image denoising," in *Proc. IEEE Conf. Comput. Vis.*, 2015, pp. 244–252.
- [43] D. L. Donoho, "De-noising by soft-thresholding," *IEEE Trans. Inf. Theory*, vol. 41, no. 3, pp. 613–627, 1995.
- [44] K.-q. Huang, Z.-y. Wu, G. S. Fung, and F. H. Chan, "Color image denoising with wavelet thresholding based on human visual system model," *Signal Process.: Image Commun.*, vol. 20, no. 2, pp. 115–127, 2005.
- [45] 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, 2006.
- [46] L. P. Yaroslavsky, "Local adaptive image restoration and enhancement with the use of dft and dct in a running window," in *Proc. SPIE*, vol. 2825, 1996, pp. 2–13.
- [47] A. Foi, V. Katkovnik, and K. Egiazarian, "Pointwise shape-adaptive dct for high-quality denoising and deblocking of grayscale and color images," *IEEE Trans. Image Process.*, vol. 16, no. 5, pp. 1395–1411, 2007.
- [48] J. Dai, O. C. Au, L. Fang, C. Pang, F. Zou, and J. Li, "Multichannel nonlocal means fusion for color image denoising," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 23, no. 11, pp. 1873–1886, 2013.
- [49] M. Elad and M. Aharon, "Image denoising via sparse and redundant representations over learned dictionaries," *IEEE Trans. Image Process.*, vol. 15, no. 12, pp. 3736–3745, 2006.
- [50] J. Mairal, M. Elad, and G. Sapiro, "Sparse representation for color image restoration," *IEEE Trans. Image Process.*, vol. 17, no. 1, pp. 53–69, 2007.

- [51] Y. Fu, A. Lam, I. Sato, and Y. Sato, "Adaptive spatial-spectral dictionary learning for hyperspectral image denoising," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2015, pp. 343–351.
- [52] C. Liu and W. T. Freeman, "A high-quality video denoising algorithm based on reliable motion estimation," in *Proc. Eur. Conf. Comput. Vis.*, 2010, pp. 706–719.
- [53] L. Zhang, W. Dong, D. Zhang, and G. Shi, "Two-stage image denoising by principal component analysis with local pixel grouping," *Pattern Recognit.*, vol. 43, no. 4, pp. 1531–1549, 2010.
- [54] W. Dong, G. Shi, and X. Li, "Nonlocal image restoration with bilateral variance estimation: a low-rank approach," *IEEE Trans. Image Process.*, vol. 22, no. 2, pp. 700–711, 2012.
- [55] A. Phophalia and S. K. Mitra, "3d mr image denoising using rough set and kernel pca method," *Magn. Reson. Imaging*, vol. 36, pp. 135–145, 2017.
- [56] D. Zoran and Y. Weiss, "From learning models of natural image patches to whole image restoration," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2011, pp. 479–486.
- [57] S. Hurault, T. Ehret, and P. Arias, "Eppl: an image denoising method using a gaussian mixture model learned on a large set of patches," *Image Processing On Line*, vol. 8, pp. 465–489, 2018.
- [58] 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, 2013.
- [59] J. Xu, L. Zhang, and D. Zhang, "External prior guided internal prior learning for real-world noisy image denoising," *IEEE Trans. Image Process.*, vol. 27, no. 6, pp. 2996–3010, 2018.
- [60] A. Buades, J.-L. Lisani, and M. Miladinović, "Patch-based video denoising with optical flow estimation," *IEEE Trans. Image Process.*, vol. 25, no. 6, pp. 2573–2586, 2016.
- [61] M. Rizkinia, T. Baba, K. Shirai, and M. Okuda, "Local spectral component decomposition for multi-channel image denoising," *IEEE Trans. Image Process.*, vol. 25, no. 7, pp. 3208–3218, 2016.
- [62] S. Gu, L. Zhang, W. Zuo, and X. Feng, "Weighted nuclear norm minimization with application to image denoising," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2014, pp. 2862–2869.
- [63] J. Xu, L. Zhang, D. Zhang, and X. Feng, "Multi-channel weighted nuclear norm minimization for real color image denoising," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2017, pp. 1096–1104.
- [64] J. Xu, L. Zhang, and D. Zhang, "A trilateral weighted sparse coding scheme for real-world image denoising," in *Proc. Eur. Conf. Comput. Vis.*, 2018, pp. 20–36.
- [65] P. Arias and J.-M. Morel, "Video denoising via empirical bayesian estimation of space-time patches," *J. Math. Imag. Vis.*, vol. 60, no. 1, pp. 70–93, 2018.
- [66] Y. Hou, J. Xu, M. Liu, G. Liu, L. Liu, F. Zhu, and L. Shao, "Nlh: A blind pixel-level non-local method for real-world image denoising," *IEEE Trans. Image Process.*, vol. 29, pp. 5121–5135, 2020.
- [67] R. H. Chan, C.-W. Ho, and M. Nikolova, "Salt-and-pepper noise removal by median-type noise detectors and detail-preserving regularization," *IEEE Trans. Image Process.*, vol. 14, no. 10, pp. 1479–1485, 2005.
- [68] M. Lebrun, M. Colom, and J.-M. Morel, "Multiscale image blind denoising," *IEEE Trans. Image Process.*, vol. 24, no. 10, pp. 3149–3161, 2015.
- [69] L. Zhuang and J. M. Bioucas-Dias, "Fast hyperspectral image denoising and inpainting based on low-rank and sparse representations," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 3, pp. 730–742, 2018.
- [70] J. V. Manjón, P. Coupé, A. Buades, D. L. Collins, and M. Robles, "New methods for mri denoising based on sparseness and self-similarity," *Med. Image Anal.*, vol. 16, no. 1, pp. 18–27, 2012.
- [71] P. Coupé, P. Yger, S. Prima, P. Hellier, C. Kervrann, and C. Barillot, "An optimized blockwise nonlocal means denoising filter for 3-d magnetic resonance images," *IEEE Trans. Med. Imag.*, vol. 27, no. 4, pp. 425–441, 2008.
- [72] J. V. Manjón, P. Coupé, L. Martí-Bonmatí, D. L. Collins, and M. Robles, "Adaptive non-local means denoising of mr images with spatially varying noise levels," *J. Magn. Reson. Imag.*, vol. 31, no. 1, pp. 192–203, 2010.
- [73] S. Aja-Fernández, C. Alberola-López, and C.-F. Westin, "Noise and signal estimation in magnitude mri and rician distributed images: a lmmse approach," *IEEE Trans. Image Process.*, vol. 17, no. 8, pp. 1383–1398, 2008.
- [74] J. V. Manjón, P. Coupé, and A. Buades, "Mri noise estimation and denoising using non-local pca," *Med. Image Anal.*, vol. 22, no. 1, pp. 35–47, 2015.
- [75] N. Renard, S. Bourennane, and J. Blanc-Talon, "Denoising and dimensionality reduction using multilinear tools for hyperspectral images," *IEEE Geosci. Remote. Sens. Lett.*, vol. 5, no. 2, pp. 138–142, 2008.
- [76] X. Liu, S. Bourennane, and C. Fossati, "Denoising of hyperspectral images using the parafac model and statistical performance analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 10, pp. 3717–3724, 2012.
- [77] A. Rajwade, A. Rangarajan, and A. Banerjee, "Image denoising using the higher order singular value decomposition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 4, pp. 849–862, 2012.
- [78] X. Zhang, Z. Xu, N. Jia, W. Yang, Q. Feng, W. Chen, and Y. Feng, "Denoising of 3d magnetic resonance images by using higher-order singular value decomposition," *Med. Image Anal.*, vol. 19, no. 1, pp. 75–86, 2015.
- [79] X. Zhang, J. Peng, M. Xu, W. Yang, Z. Zhang, H. Guo, W. Chen, Q. Feng, E. X. Wu, and Y. Feng, "Denoise diffusion-weighted images using higher-order singular value decomposition," *Neuroimage*, vol. 156, pp. 128–145, 2017.
- [80] Y. Peng, D. Meng, Z. Xu, C. Gao, Y. Yang, and B. Zhang, "Decomposable nonlocal tensor dictionary learning for multispectral image denoising," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2014, pp. 2949–2956.
- [81] Z. Zhang and S. Aeron, "Denoising and completion of 3d data via multidimensional dictionary learning," in *Proc. 25th Int. Joint Conf. Artif. Intell.*, 2016, pp. 2371–2377.
- [82] 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, 2015.
- [83] Q. Xie, Q. Zhao, D. Meng, Z. Xu, S. Gu, W. Zuo, and L. Zhang, "Multispectral images denoising by intrinsic tensor sparsity regularization," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 1692–1700.
- [84] Y. Chang, L. Yan, and S. Zhong, "Hyper-laplacian regularized unidirectional low-rank tensor recovery for multispectral image denoising," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 4260–4268.
- [85] W. He, H. Zhang, H. Shen, and L. Zhang, "Hyperspectral image denoising using local low-rank matrix recovery and global spatiotemporal total variation," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 3, pp. 713–729, 2018.
- [86] Y. Wu, L. Fang, and S. Li, "Weighted tensor rank-1 decomposition for nonlocal image denoising," *IEEE Trans. Image Process.*, vol. 28, no. 6, pp. 2719–2730, 2018.
- [87] H. Lv and R. Wang, "Denoising 3d magnetic resonance images based on low-rank tensor approximation with adaptive multi-rank estimation," *IEEE Access*, vol. 7, pp. 85995–86003, 2019.
- [88] Z. Kong, L. Han, X. Liu, and X. Yang, "A new 4-d nonlocal transform-domain filter for 3-d magnetic resonance images denoising," *IEEE Trans. Med. Imag.*, vol. 37, no. 4, pp. 941–954, 2017.
- [89] W. He, Q. Yao, C. Li, N. Yokoya, and Q. Zhao, "Non-local meets global: An integrated paradigm for hyperspectral denoising," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 6861–6870.
- [90] X. Gong, W. Chen, and J. Chen, "A low-rank tensor dictionary learning method for hyperspectral image denoising," *IEEE Trans. Signal Process.*, vol. 68, pp. 1168–1180, 2020.
- [91] W. Dong, G. Li, G. Shi, X. Li, and Y. Ma, "Low-rank tensor approximation with laplacian scale mixture modeling for multiframe image denoising," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2015, pp. 442–449.
- [92] A. S. Rubel, V. V. Lukin, and K. O. Egiazarian, "Metric performance in similar blocks search and their use in collaborative 3d filtering of grayscale images," in *Image Processing: Algorithms and Systems XII*, vol. 9019, 2014.
- [93] A. Foi and G. Boracchi, "Foveated nonlocal self-similarity," *Int. J. Comput. Vis.*, vol. 120, no. 1, pp. 78–110, 2016.
- [94] T. Ehret, P. Arias, and J.-M. Morel, "Global patch search boosts video denoising," in *Proc. Int. Conf. Comput. Vis. Theory Appl.*, vol. 5, 2017, pp. 124–134.
- [95] Y. Mkinen, L. Azzari, and A. Foi, "Collaborative filtering of correlated noise: Exact transform-domain variance for improved shrinkage and patch matching," *IEEE Trans. Image Process.*, 2020.

- [96] J. Liu, P. Musialski, P. Wonka, and J. Ye, "Tensor completion for estimating missing values in visual data," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 1, pp. 208–220, 2012.
- [97] L. R. Tucker, "Some mathematical notes on three-mode factor analysis," *Psychometrika*, vol. 31, no. 3, pp. 279–311, 1966.
- [98] L. De Lathauwer, B. De Moor, and J. Vandewalle, "A multilinear singular value decomposition," *SIAM J. Matrix Anal. Appl.*, vol. 21, no. 4, pp. 1253–1278, 2000.
- [99] M. E. Kilmer and C. D. Martin, "Factorization strategies for third-order tensors," *Linear Algebra Appl.*, vol. 435, no. 3, pp. 641–658, 2011.
- [100] M. E. Kilmer, K. Braman, N. Hao, and R. C. Hoover, "Third-order tensors as operators on matrices: A theoretical and computational framework with applications in imaging," *SIAM J. Matrix Anal. Appl.*, vol. 34, no. 1, pp. 148–172, 2013.
- [101] D. L. Donoho and J. M. Johnstone, "Ideal spatial adaptation by wavelet shrinkage," *biometrika*, vol. 81, no. 3, pp. 425–455, 1994.
- [102] D. Muti, S. Bourennane, and J. Marot, "Lower-rank tensor approximation and multiway filtering," *SIAM J. Matrix Anal. Appl.*, vol. 30, no. 3, pp. 1172–1204, 2008.
- [103] I. V. Oseledets, "Tensor-train decomposition," *SIAM J. Sci. Comput.*, vol. 33, no. 5, pp. 2295–2317, 2011.
- [104] J. A. Bengua, H. N. Phien, H. D. Tuan, and M. N. Do, "Efficient tensor completion for color image and video recovery: Low-rank tensor train," *IEEE Trans. Image Process.*, vol. 26, no. 5, pp. 2466–2479, 2017.
- [105] H. C. Burger, C. J. Schuler, and S. Harmeling, "Image denoising: Can plain neural networks compete with bm3d?" in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2012, pp. 2392–2399.
- [106] Y. Chen and T. Pock, "Trainable nonlinear reaction diffusion: A flexible framework for fast and effective image restoration," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 6, pp. 1256–1272, 2016.
- [107] 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, 2017.
- [108] D. Jiang, W. Dou, L. Vosters, X. Xu, Y. Sun, and T. Tan, "Denoising of 3d magnetic resonance images with multi-channel residual learning of convolutional neural network," *Jpn. J. Radiol.*, vol. 36, no. 9, pp. 566–574, 2018.
- [109] S. Lefkimiou, "Non-local color image denoising with convolutional neural networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 3587–3596.
- [110] ——, "Universal denoising networks : A novel cnn architecture for image denoising," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 3204–3213.
- [111] K. Zhang, W. Zuo, and L. Zhang, "Ffdnet: Toward a fast and flexible solution for cnn-based image denoising," *IEEE Trans. Image Process.*, vol. 27, no. 9, pp. 4608–4622, 2018.
- [112] 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, 2018.
- [113] 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, 2019.
- [114] M. Claus and J. van Gemert, "Videnn: Deep blind video denoising," in *Proc. Conf. Comput. Vis. Pattern Recognit. Workshops*, 2019, pp. 1–10.
- [115] M. Chang, Q. Li, H. Feng, and Z. Xu, "Spatial-adaptive network for single image denoising," *arXiv preprint arXiv:2001.10291*, 2020.
- [116] G. Vaksman, M. Elad, and P. Milanfar, "Lidia: Lightweight learned image denoising with instance adaptation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, June 2020.
- [117] Y. Zhao, D. Zhai, J. Jiang, and X. Liu, "Adrn: Attention-based deep residual network for hyperspectral image denoising," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, 2020, pp. 2668–2672.
- [118] Y. Chang, L. Yan, H. Fang, S. Zhong, and W. Liao, "Hsi-denet: Hyperspectral image restoration via convolutional neural network," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 2, pp. 667–682, 2018.
- [119] J. Lehtinen, J. Munkberg, J. Hasselgren, S. Laine, T. Karras, M. Aittala, and T. Aila, "Noise2noise: Learning image restoration without clean data," in *Proc. Int. Conf. Mach. Learn.*, 2018, pp. 2965–2974.
- [120] Z. Yue, H. Yong, Q. Zhao, D. Meng, and L. Zhang, "Variational denoising network: Toward blind noise modeling and removal," in *Proc. Advances Neural Inf. Process. Syst.*, 2019, pp. 1690–1701.
- [121] S. Guo, Z. Yan, K. Zhang, W. Zuo, and L. Zhang, "Toward convolutional blind denoising of real photographs," in *Proc. Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 1712–1722.
- [122] J. V. Manjón and P. Coupé, "Mri denoising using deep learning," in *Patch-Based Techniques in Medical Imaging*, 2018, pp. 12–19.
- [123] A. Abbasi, A. Monadjemi, L. Fang, H. Rabbani, and Y. Zhang, "Three-dimensional optical coherence tomography image denoising through multi-input fully-convolutional networks," *Comput. Biol. Med.*, vol. 108, pp. 1–8, 2019.
- [124] S. Yu, B. Park, and J. Jeong, "Deep iterative down-up cnn for image denoising," in *Proc. Conf. Comput. Vis. Pattern Recognit. Workshops*, 2019.
- [125] Y. Song, Y. Zhu, and X. Du, "Dynamic residual dense network for image denoising," *Sensors*, vol. 19, no. 17, p. 3809, 2019.
- [126] S. Gu, Y. Li, L. V. Gool, and R. Timofte, "Self-guided network for fast image denoising," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2019, pp. 2511–2520.
- [127] C. Chen, Z. Xiong, X. Tian, Z.-J. Zha, and F. Wu, "Real-world image denoising with deep boosting," *IEEE Trans. Pattern Anal. Mach. Intell.*, 2019.
- [128] S. Anwar and N. Barnes, "Real image denoising with feature attention," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2019, pp. 3155–3164.
- [129] Y. Kim, J. W. Soh, G. Y. Park, and N. I. Cho, "Transfer learning from synthetic to real-noise denoising with adaptive instance normalization," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2020, pp. 3482–3492.
- [130] Y. Quan, M. Chen, T. Pang, and H. Ji, "Self2self with dropout: Learning self-supervised denoising from single image," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2020, pp. 1890–1898.
- [131] K. Wei, Y. Fu, and H. Huang, "3-d quasi-recurrent neural network for hyperspectral image denoising," *IEEE Trans. Neural Netw. Learn. Syst.*, 2020.
- [132] S. W. Zamir, A. Arora, S. Khan, M. Hayat, F. S. Khan, M.-H. Yang, and L. Shao, "Cycleisp: Real image restoration via improved data synthesis," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2020, pp. 2696–2705.
- [133] Z. Yue, Q. Zhao, L. Zhang, and D. Meng, "Dual adversarial network: Toward real-world noise removal and noise generation," *arXiv preprint arXiv:2007.05946*, 2020.
- [134] S. W. Zamir, A. Arora, S. Khan, M. Hayat, F. S. Khan, M.-H. Yang, and L. Shao, "Learning enriched features for real image restoration and enhancement," in *Proc. Eur. Conf. Comput. Vis.*, 2020.
- [135] D. Valsesia, G. Fracastoro, and E. Magli, "Deep graph-convolutional image denoising," *IEEE Trans. Image Process.*, vol. 29, pp. 8226–8237, 2020.
- [136] J. Chen, J. Chen, H. Chao, and M. Yang, "Image blind denoising with generative adversarial network based noise modeling," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 3155–3164.
- [137] H. Yan, X. Chen, V. Y. Tan, W. Yang, J. Wu, and J. Feng, "Unsupervised image noise modeling with self-consistent gan," *arXiv preprint arXiv:1906.05762*, 2019.
- [138] R. A. Yeh, T. Y. Lim, C. Chen, A. G. Schwing, M. Hasegawa-Johnson, and M. Do, "Image restoration with deep generative models," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, 2018, pp. 6772–6776.
- [139] K. Lin, T. H. Li, S. Liu, and G. Li, "Real photographs denoising with noise domain adaptation and attentive generative adversarial network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, June 2019.
- [140] A. Abdelhamed, R. Timofte, M. S. Brown, S. Yu, B. Park, J. Jeong, S. W. Jung, D. W. Kim, J. R. Chung, J. Liu *et al.*, "Ntire 2019 challenge on real image denoising: Methods and results," in *Proc. Conf. Comput. Vis. Pattern Recognit. Workshops*, 2019, pp. 2197–2210.
- [141] A. Abdelhamed, S. Lin, and M. S. Brown, "A high-quality denoising dataset for smartphone cameras," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 1692–1700.
- [142] D. Ulyanov, A. Vedaldi, and V. Lempitsky, "Deep image prior," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 9446–9454.

- [143] Y. Zhou, R. Chellappa, and B. Jenkins, "A novel approach to image restoration based on a neural network," in *IEEE International Conference on Neural Networks*, vol. 4, 1987, pp. 269–276.
- [144] Y.-W. Chiang and B. Sullivan, "Multi-frame image restoration using a neural network," in *Proc. Midwest Symp. Circuits Syst.* IEEE, 1989, pp. 744–747.
- [145] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.
- [146] S. Ji, W. Xu, M. Yang, and K. Yu, "3d convolutional neural networks for human action recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 1, pp. 221–231, 2012.
- [147] P. Moeskops, M. A. Viergever, A. M. Mendrik, L. S. De Vries, M. J. Binders, and I. Isgum, "Automatic segmentation of mr brain images with a convolutional neural network," *IEEE Trans. Med. Imag.*, vol. 35, no. 5, pp. 1252–1261, 2016.
- [148] W. Shi, J. Caballero, F. Huszár, J. Totz, A. P. Aitken, R. Bishop, D. Rueckert, and Z. Wang, "Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 1874–1883.
- [149] L. Windrim, A. Melkumyan, R. J. Murphy, A. Chlingaryan, and R. Ramakrishnan, "Pretraining for hyperspectral convolutional neural network classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 5, pp. 2798–2810, 2018.
- [150] V. Jain and S. Seung, "Natural image denoising with convolutional networks," in *Proc. Advances Neural Inf. Process. Syst.*, 2009, pp. 769–776.
- [151] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *Proc. Int. Conf. Mach. Learn.*, 2015, pp. 448–456.
- [152] V. Nair and G. E. Hinton, "Rectified linear units improve restricted boltzmann machines," in *Proc. Int. Conf. Mach. Learn.*, 2010, pp. 807–814.
- [153] A. Radford, L. Metz, and S. Chintala, "Unsupervised representation learning with deep convolutional generative adversarial networks," *arXiv preprint arXiv:1511.06434*, 2015.
- [154] P. Chatterjee and P. Milanfar, "Is denoising dead?" *IEEE Trans. Image Process.*, vol. 19, no. 4, pp. 895–911, 2009.
- [155] X. Zhang and R. Wu, "Fast depth image denoising and enhancement using a deep convolutional network," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, 2016, pp. 2499–2503.
- [156] S. Nam, Y. Hwang, Y. Matsushita, and S. Joo Kim, "A holistic approach to cross-channel image noise modeling and its application to image denoising," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 1683–1691.
- [157] J. Anaya and A. Barbu, "Renoir—a dataset for real low-light image noise reduction," *J. Vis. Commun. Image Represent.*, vol. 51, pp. 144–154, 2018.
- [158] J. Xu, H. Li, Z. Liang, D. Zhang, and L. Zhang, "Real-world noisy image denoising: A new benchmark," *arXiv preprint arXiv:1804.02603*, 2018.
- [159] F. Perazzi, J. Pont-Tuset, B. McWilliams, L. Van Gool, M. Gross, and A. Sorkine-Hornung, "A benchmark dataset and evaluation methodology for video object segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 724–732.
- [160] H. Yue, C. Cao, L. Liao, R. Chu, and J. Yang, "Supervised raw video denoising with a benchmark dataset on dynamic scenes," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2020, pp. 2301–2310.
- [161] F. Yasuma, T. Mitsunaga, D. Iso, and S. Nayar, "Generalized Assorted Pixel Camera: Post-Capture Control of Resolution, Dynamic Range and Spectrum," Tech. Rep., Nov 2008.
- [162] B. Arad and O. Ben-Shahar, "Sparse recovery of hyperspectral signal from natural rgb images," in *Proc. Eur. Conf. Comput. Vis.*, 2016, pp. 19–34.
- [163] A. Chakrabarti and T. Zickler, "Statistics of real-world hyperspectral images," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2011, pp. 193–200.
- [164] C. A. Cocosco, V. Kollokian, R. K.-S. Kwan, G. B. Pike, and A. C. Evans, "Brainweb: Online interface to a 3d mri simulated brain database," in *NeuroImage*, 1997.
- [165] D. S. Marcus, T. H. Wang, J. Parker, J. G. Csernansky, J. C. Morris, and R. L. Buckner, "Open access series of imaging studies (oasis): cross-sectional mri data in young, middle aged, nondemented, and demented older adults," *J. Cogn. Neurosci.*, vol. 19, no. 9, pp. 1498–1507, 2007.
- [166] Kodak, "Kodak gallery dataset," [Online]. Available: <http://r0k.us/graphics/kodak>.
- [167] Derf, "Derfs test media collection," [Online]. Available: <https://media.xiph.org/video/derf>.
- [168] K. Wei, Y. Fu, J. Yang, and H. Huang, "A physics-based noise formation model for extreme low-light raw denoising," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2020, pp. 2758–2767.
- [169] J. M. Amigo and S. Grassi, "Configuration of hyperspectral and multispectral imaging systems," in *Data Handling in Science and Technology*, 2020, vol. 32, pp. 17–34.
- [170] D. S. Elson, "Interventional imaging: Biophotonics," in *Handbook of Medical Image Computing and Computer Assisted Intervention*. Elsevier, 2020, pp. 747–775.
- [171] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, 2004.
- [172] Y. Fang, H. Zhu, K. Ma, Z. Wang, and S. Li, "Perceptual evaluation for multi-exposure image fusion of dynamic scenes," *IEEE Trans. Image Process.*, vol. 29, pp. 1127–1138, 2019.
- [173] R. H. Yuhas, J. W. Boardman, and A. F. Goetz, "Determination of semi-arid landscape endmembers and seasonal trends using convex geometry spectral unmixing techniques," *ratio*, vol. 4, p. 22, 1990.
- [174] L. Wald, *Data fusion: definitions and architectures: fusion of images of different spatial resolutions*. Les Presses de l'Ecole Mines, 2002.
- [175] A. Foi, "Noise estimation and removal in mr imaging: The variance-stabilization approach," in *Proc. IEEE Int. Symp. Biomed. Imag., Nano Macro*, 2011, pp. 1809–1814.
- [176] T. Rabie, "Robust estimation approach for blind denoising," *IEEE Trans. Image Process.*, vol. 14, no. 11, pp. 1755–1765, 2005.
- [177] A. Majumdar, "Blind denoising autoencoder," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 30, no. 1, pp. 312–317, 2018.
- [178] J. Chen, J. Benesty, Y. Huang, and S. Doclo, "New insights into the noise reduction wiener filter," *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 14, no. 4, pp. 1218–1234, 2006.
- [179] E. Luo, S. Pan, and T. Nguyen, "Generalized non-local means for iterative denoising," in *Proc. 20th Eur. Signal Process. Conf.*, 2012, pp. 260–264.
- [180] T. Tirer and R. Giryes, "Image restoration by iterative denoising and backward projections," *IEEE Trans. Image Process.*, vol. 28, no. 3, pp. 1220–1234, 2018.
- [181] M. Zontak, I. Mosseri, and M. Irani, "Separating signal from noise using patch recurrence across scales," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2013, pp. 1195–1202.
- [182] C. Dong, C. C. Loy, K. He, and X. Tang, "Image super-resolution using deep convolutional networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 2, pp. 295–307, 2015.
- [183] Z. Wang, J. Chen, and S. C. Hoi, "Deep learning for image super-resolution: A survey," *IEEE Trans. Pattern Anal. Mach. Intell.*, 2020.
- [184] S. Abramov, V. Lukin, O. Rubel, and K. Egiazarian, "Prediction of performance of 2d dct-based filter and adaptive selection of its parameters," *Electronic Imaging*, vol. 2020, no. 9, 2020.
- [185] Q. Wu, W. Ren, and X. Cao, "Learning interleaved cascade of shrinkage fields for joint image dehazing and denoising," *IEEE Trans. Image Process.*, vol. 29, pp. 1788–1801, 2019.
- [186] L. Liu, X. Jia, J. Liu, and Q. Tian, "Joint demosaicing and denoising with self guidance," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2020, pp. 2240–2249.
- [187] D. Liu, B. Wen, J. Jiao, X. Liu, Z. Wang, and T. S. Huang, "Connecting image denoising and high-level vision tasks via deep learning," *IEEE Trans. Image Process.*, vol. 29, pp. 3695–3706, 2020.
- [188] J. Kamm and J. G. Nagy, "Kronecker product and svd approximations in image restoration," *Linear Algebra Appl.*, vol. 284, no. 1-3, pp. 177–192, 1998.
- [189] L. De Lathauwer, B. De Moor, and J. Vandewalle, "On the best rank-1 and rank-(r1, r2,..., rn) approximation of higher-order tensors," *SIAM J. Matrix Anal. Appl.*, vol. 21, no. 4, pp. 1324–1342, 2000.
- [190] A. Cichocki, D. Mandic, L. De Lathauwer, G. Zhou, Q. Zhao, C. Caiafa, and H. A. Phan, "Tensor decompositions for signal processing applications: From two-way to multiway component analysis," *IEEE Sig. Process. Mag.*, vol. 32, no. 2, pp. 145–163, 2015.
- [191] A. Zhang and D. Xia, "Tensor svd: Statistical and computational limits," *IEEE Trans. Inf. Theory*, vol. 64, no. 11, pp. 7311–7338, 2018.