

000  
001  
002054  
055  
056

# External Patch Group Prior Guided Internal Orthogonal Dictionary Learning for Real Image Denoising

057  
058  
059003  
004  
005  
006  
007060  
061  
062  
063

Anonymous CVPR submission

008  
009  
010  
011  
012064  
065  
066  
067

Paper ID \*\*\*\*

013  
014  
015068  
069  
070

## Abstract

016  
017  
018  
019  
020071  
072  
073  
074

For image denoising problem, the external and internal priors are playing key roles in many different methods. External priors learn from external images to restore noisy images while internal ones exploit priors of given images for denoising. The external priors are more generative and efficient on recovering structures existing in most images while the internal priors are more adaptive on recovering details existed in given noisy images. In this paper, we propose to employ the external patch group prior of images to guide the clustering of internal patch groups, and develop an external dictionary guided internal orthogonal dictionary learning algorithm for real image denoising. The internal orthogonal dictionary learning process has closed-form solutions and hence very efficient for online denoising. The experiments on standard datasets demonstrate that, that the proposed method achieves better performance than other state-of-the-art methods on real image denoising.

033  
034  
035075  
076  
077

## 1. Introduction

036  
037  
038  
039  
040078  
079  
080  
081  
082

Most vision systems, such as medical imaging and surveillance, need accurate feature extraction from high-quality images. The camera sensors and outdoor low light conditions will unavoidably bring noise to the captured images. The impact is that the image details will be lost or hardly visible. As a result, image denoising is an essential procedure for the reliability of these vision systems. In the research area, image denoising is also an ideal platform for testing natural image models and provides high-quality images for other computer vision tasks such as image registration, segmentation, and pattern recognition, etc.

041  
042  
043  
044  
045  
046  
047  
048  
049  
050  
051  
052  
053083  
084  
085  
086  
087  
088  
089  
090  
091  
092  
093  
094  
095  
096  
097  
098  
099  
100  
101  
102  
103  
104  
105  
106  
107

For several decades, there emerge numerous image denoising methods [1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11], and all of them focus mainly on dealing with additive white Gaussian noise (AWGN). In real world, the cameras will undertake high ISO settings for high-speed shots on actions, long exposure for low light on night shots, etc. Under these

situations, the noise is generated in a complex form and also been changed during the in-camera imaging pipeline [12, 13]. Therefore, the noise in real images are much more complex than Gaussian [13, 14]. It depends on camera series, brands, as well as the settings (ISO, shutter speed, and aperture, etc). The models designed for AWGN would become much less effective on real noisy images.

In the last decade, the methods of [15, 16, 17, 18, 19, 20, 13] are developed to deal with real noisy images. Almost all these methods employ a two-stage framework: estimating the parameters of the assumed noise model (usually Gaussian) and performing denoising with the help of the noise modeling and estimation in the first stage. However, the Gaussian assumption is inflexible in describing the complex noise on real noisy images [17]. Although the mixture of Gaussians (MoG) model is possible to approximate any noise distribution [21], estimating its parameters is time consuming via nonparametric Bayesian techniques [20]. To evaluate the performance of these methods on dealing with complex real noise, we apply these methods, with corresponding default parameters, on a real noisy image provided in [13]. The testing image is captured by a Nikon D800 camera when ISO is 3200. The "ground truth" image is also provided with which we can calculate objective measurements such as PSNR and SSIM [22]. The denoised images are listed in Figure 1, from which we can see that these methods either remove the noise or oversmooth the complex details in real noisy image.

The above mentioned methods can be categorized into external methods which learn priors from external images to recover noisy images, and internal ones which exploit priors of given images for denoising. The external priors in natural images are free of the high correlation between noise and signals in real noisy images, while the internal prior is adaptive to the image and can recover better the latent clean image. Combining the priors of external clean images and adaptively of internal testing images can naturally improve the performance of denoising methods, especially on real noisy images. Based on these observations, in this paper, we propose to employ the external patch group prior [10]

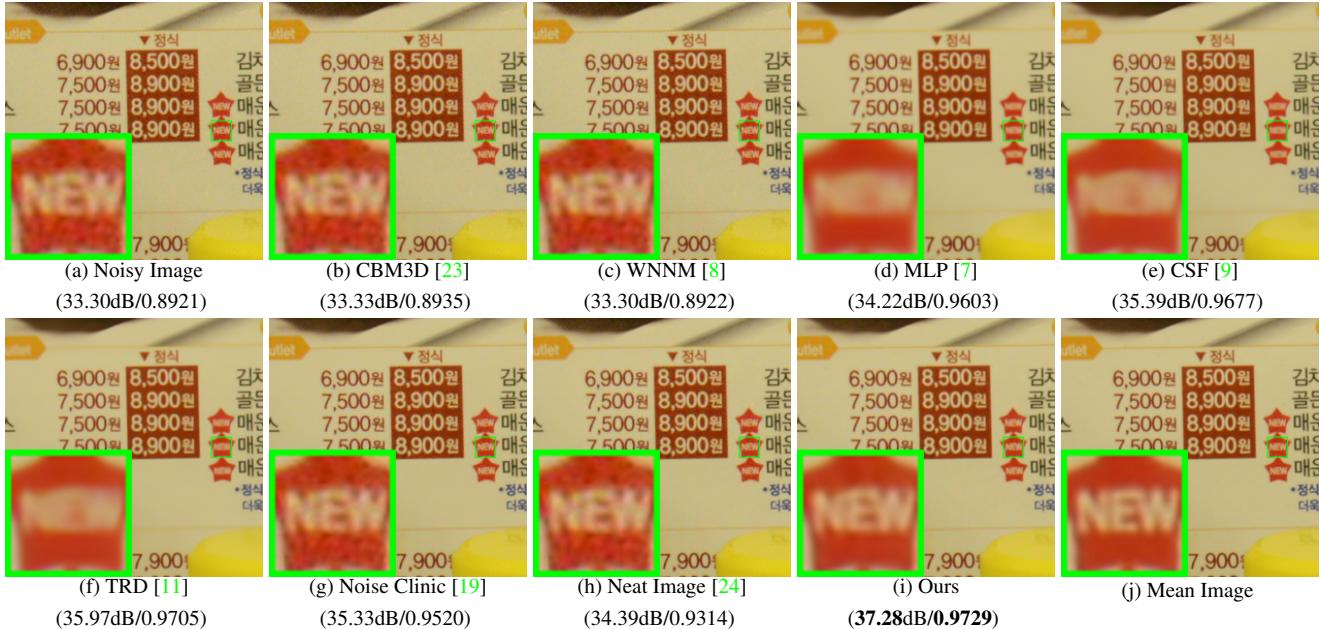


Figure 1. Denoised images of the real noisy image "Nikon D800 ISO 3200 A3" from [13] by different methods. The images are better viewed by zooming in on screen.

of natural clean images to guide the clustering of internal patch groups in given noisy image, and develop an external prior guided internal orthogonal dictionary learning (DL) algorithm for real image denoising. The internal orthogonal DL process includes two alternating stages: updating sparse coefficients and updating orthogonal dictionary. Both of the two stages have closed-form solutions. Hence, our internal DL process is very efficient for online internal denoising. Through comprehensive experiments on real noisy images captured by different cameras and settings, we demonstrate that the proposed method achieves better performance on real image denoising

## 1.1. Our Contributions

The contributions of this paper are summarized as follows:

- We propose a novel dictionary learning method which employs the external prior guided the internal orthogonal dictionary learning for real image denoising. Both the external prior and internal prior are performed on patch groups instead of patches.
- The internal orthogonal dictionary learning are alternating iterative solved with closed-form solutions. The learned orthogonal dictionary are very efficient in both learning and denoising stages.
- We achieve much better performance on the real image denoising problem than other competing methods in terms of visual quality, PSNR, and SSIM.

The rest of this paper will be summarized as follows: in Section 2, we will introduce the related work; in Section 3, we will introduce the proposed external prior guided internal orthogonal dictionary learning algorithm for real image denoising; in Section 4, we will demonstrate the extensive experiments on two standard dataset; we will conclude our paper and give our future work in Section 5.

## 2. Related Work

### 2.1. Patch Group Prior of Natural Images

The Patch Group (PG) prior [10] is proposed to directly model the non-local self similar (NSS) property of natural images. The NSS property is commonly used in image restoration tasks [1, 4, 5, 8, 10]. The PG prior largely reduces the space of images to be modeled when compared to the patch prior [6]. In [10], only the PGs of clean natural images is utilized, while the PGs of noisy input images are ignored. In this paper, we make use of PGs both from external clean images and internal given real noisy image for better denoising performance.

### 2.2. Internal v.s. External Dictionary Learning

There are two categories of dictionary learning methods in image denoising. One category of methods [6, 10] learns dictionary from external clean images to reconstruct the given noisy images, while the other category of methods [25, 33] directly learns adaptive dictionary from the given noisy image to recover the latent clean image. The two categories of methods are very successful in removing Gaussian noise. However, the noise in real images is generated mostly

from the camera sensors, which is highly complex and signal dependent [13]. Besides, the noise in real images has fixed patterns from several main sources [14]. Therefore, we can hardly separate the complex noise from the signals without the help of external (correct) information of natural clean images. In this paper, our goal is to employ the external patch group prior to guide the internal noisy PGs to be clustered into correct subspace, and the learning process of internal orthogonal dictionary.

### 2.3. Real Image Denoising

In the last decade, there are many methods [15, 16, 17, 18, 19, 20, 13] proposed for real image denoising problem. In the seminar work of BLS-GSM [30] for real image denoising, Portilla et al. proposed to use scale mixture of Gaussian in overcomplete oriented pyramids to estimate the latent clean images. In [15], Portilla proposed to use a correlated Gaussian model for noise estimation of each wavelet subband. The work of Rabie [16] modeled the noisy pixels as outliers which are removed via Lorentzian robust estimator [31]. Liu et al. [17] proposed to use ‘noise level function’ (NLF) to estimate the noise and then use Gaussian conditional random field to obtain the latent clean image. Gong et al. [18] models the noise by mixed  $\ell_1$  and  $\ell_2$  norms and remove the noise by sparsity prior in the wavelet transform domain. Later, Lebrun et al. proposed a multi-scale denoising algorithm called ‘Noise Clinic’ [19]. This method generalizes the NL-Bayes model [32] to deal with blind noise and achieves state-of-the-art performance. Recently, Zhu et al. proposed a Bayesian model [20] which approximates and remove the noise via Low-Rank Mixture of Gaussian (LR-MoG) model. The commercial software Neat Image [24] first estimate the parameters of noise via a large flat area and then filter the noise accordingly.

## 3. External Patch Group Prior Guided Internal Orthogonal Dictionary Learning

In this section, we formulate the framework of external Patch Group prior guided internal subspace learning. We first introduce the patch group prior leaning on clean natural RGB images. Then we formulate the external guided internal subspace learning. Finally, we discuss the differences between external subspaces and the corresponding internal subspace.

### 3.1. External Patch Group Prior Learning

Natural images often demonstrate repetitive patterns, this nonlocal self-similarity (NSS) property is a key successful factor for many image denoising methods [1, 4, 5, 33, 8, 10] and restoration methods [ ]. In [10], the NSS property is directly learned as an external prior in a patch group manner. In this section, we formulate the Patch Group prior on natural color images.

In [10], the patch group (PG) is defined as a group of similar patches to the local patch. The patch group mean is detracted, and hence different groups patches can share similar PGs. Therefore the space to be modeled is largely reduced. In this work, we extract PGs from RGB images. Each patch is of size  $p \times p \times 3$ . For each local patch, we search its similar patches around it through the Euclidean distance in a local window of size  $W \times W$ . The PG is denoted by  $\{\mathbf{x}_m\}_{m=1}^M$ , where  $\mathbf{x}_m \in \mathbb{R}^{3p^2 \times 1}$  is a color image patch vector. The mean vector of this PG is  $\boldsymbol{\mu} = \frac{1}{M} \sum_{m=1}^M \mathbf{x}_m$ , and  $\bar{\mathbf{x}}_m = \mathbf{x}_m - \boldsymbol{\mu}$  is the group mean subtracted patch vector. The PG is defined as  $\bar{\mathbf{X}} \triangleq \{\bar{\mathbf{x}}_m\}, m = 1, \dots, M$ , and it represent the external NSS prior on color images. Assume we have extracted  $N$  PGs from a given set of natural images, and the  $n$ -th PG is defined as  $\bar{\mathbf{X}}_n \triangleq \{\bar{\mathbf{x}}_{n,m}\}_{m=1}^M, n = 1, \dots, N$ . We employ the patch group based Gaussian Mixture Model (PG-GMM) for NSS prior learning. We aim to learn a set of  $K$  Gaussians  $\{\mathcal{N}(\boldsymbol{\mu}_k, \Sigma_k)\}$  from  $N$  training PGs  $\{\bar{\mathbf{X}}_n\}$ , while requiring that all the  $M$  patches  $\{\bar{\mathbf{x}}_{n,m}\}$  in PG  $\bar{\mathbf{X}}_n$  belong to the same Gaussian component and assume that the patches in the PG are independently sampled. Note that such an assumption is commonly used in patch based image modeling [3, 5]. Then, the likelihood of  $\{\bar{\mathbf{X}}_n\}$  can be calculated as

$$P(\bar{\mathbf{X}}_n) = \sum_{k=1}^K \pi_k \prod_{m=1}^M \mathcal{N}(\bar{\mathbf{x}}_{n,m} | \boldsymbol{\mu}_k, \Sigma_k). \quad (1)$$

By assuming that all the PGs are independently sampled, the overall objective log-likelihood function is

$$\ln \mathcal{L} = \sum_{n=1}^N \ln \left( \sum_{k=1}^K \pi_k \prod_{m=1}^M \mathcal{N}(\bar{\mathbf{x}}_{n,m} | \boldsymbol{\mu}_k, \Sigma_k) \right). \quad (2)$$

We maximize the above objective function for PG-GMM learning and finally obtain the GMM model with learned parameters including mixture weights  $\{\pi_k\}_{k=1}^K$ , mean vectors  $\{\boldsymbol{\mu}_k = \mathbf{0}\}_{k=1}^K$ , and covariance matrices  $\{\Sigma_k\}_{k=1}^K$ . Noted that the mean vector of each cluster is natural zeros, i.e.,  $\boldsymbol{\mu}_k = \mathbf{0}$ .

### 3.2. External Prior Guided Internal Orthogonal Dictionary Learning

Given a real noisy image, we extract noisy PGs from it and save the mean vectors of each PG for recovering. The mean subtracted PG is defined as  $\bar{\mathbf{Y}}$ . To project this PG into a most adaptive subspace, we select the most suitable Gaussian component to it from the PG-GMM trained in previous section. The selection can be done by checking the posterior probability that  $\bar{\mathbf{Y}}$  belongs to the  $k$ th Gaussian component:

$$P(k|\bar{\mathbf{Y}}) = \frac{\prod_{m=1}^M \mathcal{N}(\bar{\mathbf{y}}_m | \mathbf{0}, \Sigma_k)}{\sum_{l=1}^K \prod_{m=1}^M \mathcal{N}(\bar{\mathbf{y}}_m | \mathbf{0}, \Sigma_l)}. \quad (3)$$

324 Since the noise on real images are mostly small when compared to the signals, the covariance matrix of the  $k$ th component is still  $\Sigma_k$ . Finally, the component with the maximum A-posteriori (MAP) probability  $\ln P(k|\bar{\mathbf{Y}})$  is selected as the most suitable subspace for  $\bar{\mathbf{Y}}$ .

325 Though each PG has been projected into its most suitable  
 326 subspace, the pre-learned subspace is still too general to  
 327 represent the noisy PG extracted from the real noisy image.  
 328 That is, the noisy PGs projected into one cluster can still  
 329 constisted a subspace which is of lower dimensions than the  
 330 subspace pre-learned from the external PGs. This can be  
 331 demonstrated by compare the distribution of external PGs  
 332 and internal PGs in the same clusters. We randomly select  
 333 one cluster, and collect the celan PGs extracted from external  
 334 dataset (Kodak 24 images) and the niosy PGs from the  
 335 testing image. Since the original PGs are of  $3p^2$  dimensions,  
 336 we apply PCA to project the PGs into 2 dimensions for better  
 337 visualization. The results is shown in Figure ??, from  
 338 which we can see clearly that the projected PGs are mainly  
 339 in a smaller region of the external PGs, which proves that  
 340 the internal PGs are only consisted a subspace in a lower  
 341 dimension than the PGs collected from external subspace.  
 342 To better and adaptively charactering the internal PGs from  
 343 the testing image, we need learn a more specific dictionary  
 344 for noisy PGs assigned into each cluster. For notation sim-  
 345 plicity, we ignore the index of subspace  $k$ . The internal PGs  
 346  $\mathbf{Y}$  form a subspace which can be obtained by singular value  
 347 decomposition (SVD),  
 348

$$\begin{aligned} & \min_{\mathbf{D}_i \in \mathbb{R}^{3p^2 \times r}, \mathbf{A} \in \mathbb{R}^{3p^2 \times MN}} \|\mathbf{Y} - [\mathbf{D}_e \mathbf{D}_i] \mathbf{A}\|_F^2 + \lambda \|\mathbf{A}\|_1 \\ & \text{s.t. } \mathbf{D}_i^T \mathbf{D}_i = \mathbf{I}_r, \mathbf{D}_e^T \mathbf{D}_i = \mathbf{0}, \end{aligned} \quad (4)$$

349 The singular vectors capture the statistical structures of NSS  
 350 variations in natural images, while the singular values in  $\mathbf{S}$   
 351 represent the significance of these singular vectors. Fig. 4  
 352 shows the singular vectors for one Gaussian component.

### 3.3. Optimization with Closed-form Solution

363 Similar to the K-SVD [3], we employ an alternating it-  
 364 erative framework to solve the optimization problem 4. In  
 365 fact, we initialize the orthogonal dictionary as  $\mathbf{D}^{(0)}$  and for  
 366  $t = 0, 1, \dots, T - 1$ , alternatively do

367 **Updating Sparse Coefficients:** given the initialization  
 368 orthogonal dictioanry  $\mathbf{D}_i^{(t)}$ , the sparce coefficients  $\mathbf{A}^{(t)}$  are  
 369 obtained via solving

$$\mathbf{A}^{(t)} := \arg \min_{\mathbf{A} \in \mathbb{R}^{3p^2 \times MN}} \|\mathbf{Y} - [\mathbf{D}_e \mathbf{D}_i^{(t)}] \mathbf{A}\|_F^2 + \lambda \|\mathbf{A}\|_1. \quad (5)$$

370 This problem has closed-form solution by  $\mathbf{A}^* =$   
 371  $T_\lambda(\hat{\mathbf{D}}^T \mathbf{Y})$ , where  $T_\lambda(\mathbf{A}) = \text{sgn}(\mathbf{A}) \odot \max(\mathbf{A}, \lambda)$  is a soft-  
 372 thresholding function.

373 **Updating Orthogonal Dictionary:** given the sparse co-  
 374 efficients  $\mathbf{A}^{(0)}$ , the sparce coefficients  $\mathbf{A}^{(t)}$  are obtained via

solving

$$\begin{aligned} \mathbf{D}_i^{(t+1)} &:= \arg \min_{\mathbf{D}_i \in \mathbb{R}^{3p^2 \times r}} \|\mathbf{Y} - [\mathbf{D}_e \mathbf{D}_i] \mathbf{A}^{(t)}\|_F^2 \\ &\text{s.t. } \mathbf{D}_i^T \mathbf{D}_i = \mathbf{I}_r, \mathbf{D}_e^T \mathbf{D}_i = \mathbf{0}, \end{aligned} \quad (6)$$

Dividing the sparse coefficients  $\mathbf{A} = [\mathbf{A}_e^T \mathbf{A}_i^T]^T$ , where  $\mathbf{A}_e$  and  $\mathbf{A}_i$  denote the coefficients over external and internal dictionary  $\mathbf{D}_e$  and  $\mathbf{D}_i$ . According to the Proposition 2.2 in [34], the problem (6) has a closed-form solution  $\mathbf{D}_i^* = \mathbf{U} \mathbf{V}^T$ , where  $\mathbf{U}$  and  $\mathbf{V}$  are the orthogonal matrices obtained by the following SVD

$$(\mathbf{I} - \mathbf{D}_e \mathbf{D}_e^T) \mathbf{Y} \mathbf{A}_i^T = \mathbf{U} \Sigma \mathbf{V}^T \quad (7)$$

With these solutions, the final obtained dictionary  $\mathbf{D} = [\mathbf{D}_e \mathbf{D}_i]$  are orthogonal ictionary. This can be proved by the following equation

$$\mathbf{D}^T \mathbf{D} = \begin{pmatrix} \mathbf{D}_e^T \\ \mathbf{D}_i^T \end{pmatrix} (\mathbf{D}_e \mathbf{D}_i) = \begin{pmatrix} \mathbf{D}_e^T \mathbf{D}_e & \mathbf{D}_e^T \mathbf{D}_i \\ \mathbf{D}_i^T \mathbf{D}_e & \mathbf{D}_i^T \mathbf{D}_i \end{pmatrix} = \mathbf{I} \quad (8)$$

### 3.4. Discussion on External Prior and Internal Orthogonal Dictionary Learning

Until now, we have divided the noisy PGs into multiple internal subspaces. Here we take a deep analysis on how the external NSS prior guide the subspace learning of internal PGs. The help are at least threefold. Firstly, through MAP in (3), the external prior guides the noisy PGs to be clustered into the correct subspaces. If we cluster the noisy PGs in an automatical way, the subspaces we learned will be highly degraded by the signal dependent noise. Secondly, the guidance of external prior for internal clustering is more efficient than directly clustering the internal noisy PGs. It only needs to calculate the MAP probability via the equation (3) while the internal clustering via GMM is time-consuming on EM algorithm [35]. Thirdly, due to the correct guidance of external prior, the strucual decompositon via SVD of each subspace is more adaptive. This will bring better denoising performance than the methods only using the external information. The *mutual incoherence*  $\mu(\mathbf{U})$  [36], which is difined as

$$\mu(\mathbf{U}) = \max_{i=j} \frac{|\mathbf{d}_i^T \mathbf{d}_j|}{\|\mathbf{d}_i\|_2 \|\mathbf{d}_j\|_2} \quad (9)$$

, is a measure of quality of dictionary.

The Internal PGs are in fact lying in the subspaces of external PG Spaces. To defend this argument, we compare the distribution of external PGs extracted from clean natural images and real noisy images. For better illumination, we randomly selected a cluster and project the original clean PGs  $\mathbf{X}$  onto a 2-D plane. This could be done via  $\mathbf{X}_p = \mathbf{U}(:, 1 : 2)^T \mathbf{X}$ , where  $\mathbf{U}$  is the singular vector matrix of that cluster. The noisy PGs  $\mathbf{Y}$  assigned in this

432 cluster is also projected into 2-D via  $\mathbf{Y}\mathbf{p} = \mathbf{U}(:, .1 : 2)^T \mathbf{Y}$ .  
 433 The Figure ?? reflects the distribution on the 2-D plane of  
 434 the projected clean PGs from external natural images and  
 435 the projected noisy PGs from internal image. We can see  
 436 that the internal noisy PGs are indeed lying in a subspace  
 437 of the external PGs. Hence, if we directly use the external  
 438 prior learned from clean PGs, the learned subspaces would  
 439 be too generative to be suitable for the testing data.  
 440

441 Through SVD, the PGs in each internal subspace can be  
 442 divided into singular vectors and singular values. The singular  
 443 vectors are the basis of the corresponding subspace  
 444 while the singular values reflect the importance of these basis.  
 445 The basis can be used as dictionary to code the noisy  
 446 PGs. And the singular values are adaptive parameters for  
 447 internal noisy PGs. We can compare the singular values of  
 448 one internal subspace and the corresponding space of external  
 449 PGs. The result is shown in Figure ???. From which  
 450 we can see that the noisy subspace often have higher values  
 451 than external space consisted of clean PGs. This gap is  
 452 clearly made of the noise and can be used for image denoising  
 453 in a natural way.

## 4. The Denoising Algorithm

### 4.1. Fast Patch Group Searching by Integral Image

454 The searching of patch groups in images is inefficient  
 455 if we search non-local similar patches to each local patch.  
 456 To speed up the searching process and make our proposed  
 457 method faster, we employ the technique of 'Summed Area  
 458 Table' [37] for efficient PG searching. The SAT permits  
 459 to evaluate the sum of pixel values in rectangular regions  
 460 of the image with four operations, regardless of the region  
 461 size. That is to say, we do not need do distance measure for  
 462 each patch. It was first proposed under the name of summed  
 463 area table[38].

### 4.2. Prior Weights for Sparse Coding

464 To remove the real noise, we employ the sparse coding  
 465 framework. And in order to be adaptive to the input image,  
 466 we employ the internal learned  $\mathbf{U}$  of each cluster as  
 467 an adaptive dictioanry to represent the structural variations  
 468 of the PGs in that cluster. Since the  $\mathbf{U}$  is orthonormal, its  
 469 *mutual incoherence* is naturally 0 and therefore better than  
 470 other redundant dictionaries.

$$\min_{\alpha} \|\bar{\mathbf{y}}_m - \mathbf{U}\alpha\|_2^2 + \sum_{i=1}^{3p^2} \lambda_i |\alpha_i|. \quad (10)$$

471 The  $i$ th entry of the regularization parameter  $\lambda_i$

$$\lambda_i = \lambda / (\mathbf{S}_i + \varepsilon), \quad (11)$$

472 where  $\varepsilon$  is a small positive number to avoid dividing by zero.  
 473 Since the dictionary  $\mathbf{U}$  is orthonormal, it is not difficult to

|  |  |
|--|--|
| <b>Alg. 1:</b> External Prior Guided Internal Orthogonal Dictionary Learning for Denoising<br><b>Input:</b> Noisy image $\mathbf{y}$ , PG-GMM model<br>1. Initialization: $\hat{\mathbf{x}}^{(0)} = \mathbf{y}, \mathbf{y}^{(0)} = \mathbf{y}$ ;<br><b>for</b> $t = 1 : IteNum$ <b>do</b><br><b>for</b> each PG $\mathbf{Y}$ <b>do</b><br>2. Calculate group mean $\mu_y$ and form PG $\bar{\mathbf{Y}}$ ;<br>3. Gaussian component selection via (3);<br><b>end for</b><br><b>for</b> each Internal Subspace <b>do</b><br>4. Internal Subspace Learning by (4);<br>5. Recover each patch in all PGs via $\hat{\mathbf{x}}_m = \mathbf{D}\hat{\alpha} + \mu_y$ ;<br><b>end for</b><br>6. Aggregate the recovered PGs of all subspaces to form the recovered image $\hat{\mathbf{x}}^{(t)}$ ;<br><b>end for</b><br><b>Output:</b> The recovered image $\hat{\mathbf{x}}^{(IteNum)}$ . | 486<br>487<br>488<br>489<br>490<br>491<br>492<br>493<br>494<br>495<br>496<br>497<br>498<br>499<br>500<br>501<br>502<br>503<br>504<br>505<br>506<br>507<br>508<br>509<br>510<br>511<br>512<br>513<br>514<br>515<br>516<br>517<br>518<br>519<br>520<br>521<br>522<br>523<br>524<br>525<br>526<br>527<br>528<br>529<br>530<br>531<br>532<br>533<br>534<br>535<br>536<br>537<br>538<br>539 |
|--|--|

find out that (4) has a closed-form solution (detailed derivation can be found in the supplementary material):

$$\hat{\alpha} = \text{sgn}(\mathbf{U}^T \bar{\mathbf{y}}_m) \odot \max(|\mathbf{U}^T \bar{\mathbf{y}}_m| - \Lambda, \mathbf{0}), \quad (12)$$

where  $\Lambda = [\lambda_1, \lambda_2, \dots, \lambda_{3p^2}]$  is the vector of regularization parameter and  $\text{sgn}(\bullet)$  is the sign function,  $\odot$  means element-wise multiplication, and  $|\mathbf{U}^T \bar{\mathbf{y}}_m|$  is the absolute value of each entry of vector  $|\mathbf{U}^T \bar{\mathbf{y}}_m|$ . The closed-form solution makes our weighted sparse coding process very efficient.

### 4.3. The Overall Algorithm

With the solution  $\hat{\alpha}$  in (7), the clean patch in a PG can be estimated as  $\hat{\mathbf{x}}_m = \mathbf{D}\hat{\alpha} + \mu_y$ . Then the clean image  $\hat{\mathbf{x}}$  can be reconstructed by aggregating all the estimated PGs. In practice, we could perform the above denoising procedures for several iterations for better denoising outputs. In iteration  $t$ , we use the iterative regularization strategy [39] to add back to the recovered image  $\hat{\mathbf{x}}^{(t-1)}$  some estimation residual in iteration  $t-1$ . The proposed denoising algorithm is summarized in Algorithm 1 (Alg. 1).

## 5. Experiments

In this section, we perform real image denoising experiments on three standard datasets. The first dataset is real noisy images with mean images as ground truths provided by [13], some samples are shown in Figure 3. The second dataset is provided by the website of Noise Clinic [19]. The third dataset is provided by the Commercial software Neat Image [24]. The second and third dataset do not have ground truth images.



Figure 2. Some testing images in the dataset [13].



Figure 3. Some cropped images of the dataset [13].

## 5.1. Implementation Details

Our proposed method contains two stages, the external prior guided internal subspace learning stage and the adaptive denoising stage. In the learning stage, there are 4 parameters: the patch size  $p$ , the number of patches in a PG  $M$ , the window size  $W$  for PG searching and the number of clusters  $K$ . We set  $p = 6$  (hence the patch size is  $6 \times 6 \times 3$ ),  $M = 10$ ,  $W = 31$ ,  $K = 32$ . We extracted about 3.6 million PGs from the Kodak PhotoCD Dataset, which includes 24 high quality color images, to train the external prior via PG-GMM. In the denoising stage, the parameter  $\lambda = 0.002$  is used to regularize the sparse term. The  $\delta$  in iterative regularization is set as  $\delta = 0.09$ .

## 5.2. Comparison on External and Internal methods

In this subsection, we compared the proposed external prior guided internal subspace learning model on real image denoising. The three methods are evaluated on the dataset provided in [13]. We calculate the PSNR, SSIM [22] and visual quality of these three methods. We also compare the speed. The PSNR and SSIM results on 60 cropped images from [13] are listed in Table 1. The images are cropped into size of  $500 \times 500$  for better illustration. We also compare the three methods on visual quality in Figure 5.2. Compare the denoised images listed in Figure 5.2 and Figure 5.2, we can see that the Offline method is better at edges, smooth regions while the Online method is good at complex textures. The reason is two folds. Firstly, the Offline method is learned on clean images and hence is better at representing edges, structures, and smooth area. The online method is influenced by the noise and hence some noise cannot be removed. Secondly, the Online method is better at recovering complex area since they could learn adaptive dictionaries for the specific area. The Offline method cannot recover the complex area since they did not learn the similar structures

Table 1. Average PSNR(dB)/SSIM results of external, internal, and guided methods on 60 cropped real noisy images in [13].

|      | Noisy  | Offline | Online | Guided        |
|------|--------|---------|--------|---------------|
| PSNR | 34.51  | 38.19   | 38.07  | <b>38.55</b>  |
| SSIM | 0.8718 | 0.9663  | 0.9625 | <b>0.9675</b> |

from the external natural clean images.

## 5.3. Comparison With other Competing Methods

We compare with previous state-of-the-art Gaussian noise removal methods such as BM3D [4], WNNM [8], MLP [7], CSF [9], and the recently proposed TRD [11]. We also compare with three competing real image denoising methods such as Noise Clinic, Neat Image, and the CCNoise method proposed recently. The popular software NeatImage which is one of the best denoising software available. All these methods need noise estimation which is very hard to perform if there is no uniform regions available in the testing image. The NeatImage will fail to perform automatical parameters settings if there is no uniform regions.<sup>1</sup>

We the competing denoising methods from various research directions on two datasets. Both the two datasets comes from the [13]. The first dataset contains 17 images of size over  $7000 \times 5000$ . Since this dataset contains repetitive contents across different images, we crop 60 small images of size  $500 \times 500$  from these 17 images in [13].

The PSNR and SSIM resluts are listed in Table 3. The number in red color and blue color means the best and second best results, respectively. From the Table 3, we can see that the external based method can already surpass largely the previous denoising methods. The improvement on PSNR over the second best method, i.e., TRD, is 0.44dB. The

## 5.4. Discussion on Parameter $\lambda$

The proposed method only has a key parameter, namely the regularization paramters  $\lambda$ . To demonstrate that the proposed method is robust to the variance of  $\lambda$ , we vary the parameter  $\lambda$  across a wide range and obtain the PSNR and SSIM results as a function of the parameter  $\lambda$ . The results is shown in Figure 8, from which we can see that the proposed method can achieve a PSNR (SSIM) over 38.5dB (0.9660) when  $\lambda$  varies from 0.0015 to 0.0025. This shows that the proposed method is indeed robust to the chosen of the paramter  $\lambda$ .

<sup>1</sup>To compare with CCNoise, we first transform the denoised images into double format.

648

649

650

651

652

653

654

655

656

657

658

659

660

661

662

663

664

665

666

667

668

669

670

671

672

673

674

675

676

677

678

679

680

681

682

683

684

685

686

687

688

689

690

691

692

693

694

695

696

697

698

699

700

701

702

703

704

705

706

707

708

709

710

711

712

713

714

715

716

717

718

719

720

721

722

723

724

725

726

727

728

729

730

731

732

733

734

735

736

737

738

739

740

741

742

743

744

745

746

747

748

749

750

751

752

753

754

755



Figure 4. Denoised images of the image "Nikon D600 ISO 3200 C1" by different methods. The images are better to be zoomed in on screen.

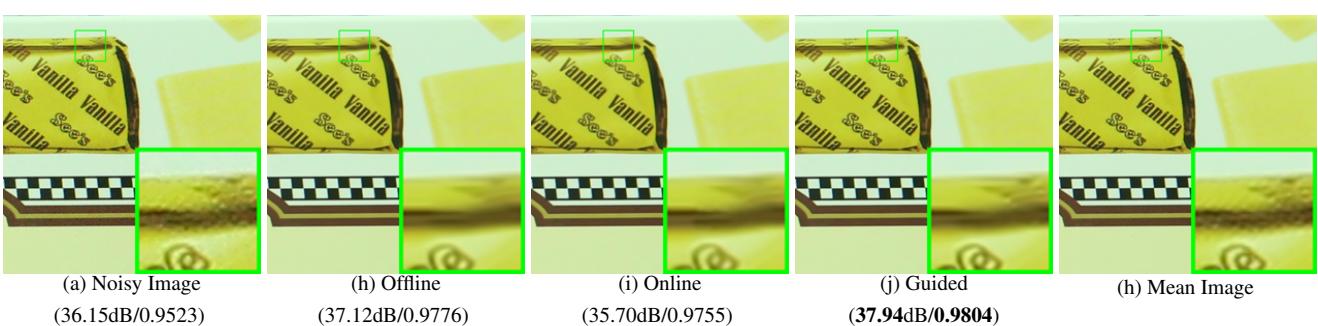


Figure 5. Denoised images of the image "Canon EOS 5D Mark3 ISO 3200 C1" by different methods. The images are better to be zoomed in on screen.

## 6. Conclusion and Future Work

In the future, we will evaluate the proposed method on other computer vision tasks such as single image super-resolution, photo-sketch synthesis, and cross-domain image recognition. Our proposed method can be improved if we use better training images, fine tune the parameters via cross-validation. We believe that our framework can be useful not just for real image denoising, but for image super-resolution, image cross-style synthesis, and recognition tasks. This will be our line of future work.

## References

- [1] A. Buades, B. Coll, and J. M. Morel. A non-local algorithm for image denoising. *CVPR*, pages 60–65, 2005. [1](#), [2](#), [3](#)
- [2] S. Roth and M. J. Black. Fields of Experts. *International Journal of Computer Vision*, 82(2):205–229, 2009. [1](#)
- [3] M. Elad and M. Aharon. Image denoising via sparse and redundant representations over learned dictionaries. *Image Processing, IEEE Transactions on*, 15(12):3736–3745, 2006. [1](#), [3](#), [4](#)
- [4] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian. Image denoising by sparse 3-D transform-domain collaborative filtering. *Image Processing, IEEE Transactions on*, 16(8):2080–2095, 2007. [1](#), [2](#), [3](#), [6](#)
- [5] J. Mairal, F. Bach, J. Ponce, G. Sapiro, and A. Zisserman. Non-local sparse models for image restoration. *ICCV*, pages 2272–2279, 2009. [1](#), [2](#), [3](#)
- [6] D. Zoran and Y. Weiss. From learning models of natural image patches to whole image restoration. *ICCV*, pages 479–486, 2011. [1](#), [2](#)
- [7] Harold C Burger, Christian J Schuler, and Stefan Harmeling. Image denoising: Can plain neural networks compete with bm3d? *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, pages 2392–2399, 2012. [1](#), [2](#), [6](#)
- [8] S. Gu, L. Zhang, W. Zuo, and X. Feng. Weighted nuclear norm minimization with application to image denoising. *CVPR*, pages 2862–2869, 2014. [1](#), [2](#), [3](#), [6](#)
- [9] U. Schmidt and S. Roth. Shrinkage fields for effective image restoration. *Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conference on*, pages 2774–2781, June 2014. [1](#), [2](#), [6](#)
- [10] J. Xu, L. Zhang, W. Zuo, D. Zhang, and X. Feng. Patch group based nonlocal self-similarity prior learning for image denoising. *2015 IEEE International Conference on Computer Vision (ICCV)*, pages 244–252, 2015. [1](#), [2](#), [3](#)
- [11] Yunjin Chen, Wei Yu, and Thomas Pock. On learning optimized reaction diffusion processes for effective image restoration. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 5261–5269, 2015. [1](#), [2](#), [6](#)
- [12] S. J. Kim, H. T. Lin, Z. Lu, S. Ssstrunk, S. Lin, and M. S. Brown. A new in-camera imaging model for color computer vision and its application. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(12):2289–2302, Dec 2012. [1](#)

756 Table 2. Average PSNR(dB) results of different methods on 60 cropped real noisy images captured in [13].  
757

|      | Noisy  | CBM3D  | WNNM   | MLP    | CSF    | TRD    | NI     | NC     | Guided | Guided2 |
|------|--------|--------|--------|--------|--------|--------|--------|--------|--------|---------|
| PSNR | 34.51  | 34.58  | 34.52  | 36.19  | 37.40  | 37.75  | 36.53  | 37.57  | 38.72  | 38.90   |
| SSIM | 0.8718 | 0.8748 | 0.8743 | 0.9470 | 0.9598 | 0.9617 | 0.9241 | 0.9514 | 0.9694 | 0.9702  |

761 Table 3. Average PSNR(dB) results of different methods on 15 cropped real noisy images used in [13].  
762

| Camera Settings                 | Noisy  | CBM3D  | WNNM   | MLP    | CSF    | TRD    | NI     | NC     | CC     | Guided2 |
|---------------------------------|--------|--------|--------|--------|--------|--------|--------|--------|--------|---------|
| Canon 5D Mark III<br>ISO = 3200 | 37.00  | 37.08  | 37.09  | 33.92  | 35.68  | 36.20  | 37.68  | 38.76  | 38.37  | 40.50   |
|                                 | 33.88  | 33.94  | 33.93  | 33.24  | 34.03  | 34.35  | 34.87  | 35.69  | 35.37  | 37.22   |
|                                 | 33.83  | 33.88  | 33.90  | 32.37  | 32.63  | 33.10  | 34.77  | 35.54  | 34.91  | 37.13   |
| Nikon D600<br>ISO = 3200        | 33.28  | 33.33  | 33.34  | 31.93  | 31.78  | 32.28  | 34.12  | 35.57  | 34.98  | 35.34   |
|                                 | 33.77  | 33.85  | 33.79  | 34.15  | 35.16  | 35.34  | 35.36  | 36.70  | 35.95  | 36.69   |
|                                 | 34.93  | 35.02  | 34.95  | 37.89  | 39.98  | 40.51  | 38.68  | 39.28  | 41.15  | 39.17   |
| Nikon D800<br>ISO = 1600        | 35.47  | 35.54  | 35.57  | 33.77  | 34.84  | 35.09  | 37.34  | 38.01  | 37.99  | 38.82   |
|                                 | 35.71  | 35.79  | 35.77  | 35.89  | 38.42  | 38.65  | 38.57  | 39.05  | 40.36  | 40.98   |
|                                 | 34.81  | 34.92  | 34.95  | 34.25  | 35.79  | 35.85  | 37.87  | 38.20  | 38.30  | 38.90   |
| Nikon D800<br>ISO = 3200        | 33.26  | 33.34  | 33.31  | 37.42  | 38.36  | 38.56  | 36.95  | 38.07  | 39.01  | 38.69   |
|                                 | 32.89  | 32.95  | 32.96  | 34.88  | 35.53  | 35.76  | 35.09  | 35.72  | 36.75  | 36.82   |
|                                 | 32.91  | 32.98  | 32.96  | 38.54  | 40.05  | 40.59  | 36.91  | 36.76  | 39.06  | 38.80   |
| Nikon D800<br>ISO = 6400        | 29.63  | 29.66  | 29.71  | 33.59  | 34.08  | 34.25  | 31.28  | 33.49  | 34.61  | 33.31   |
|                                 | 29.97  | 30.01  | 29.98  | 31.55  | 32.13  | 32.38  | 31.38  | 32.79  | 33.21  | 33.18   |
|                                 | 29.87  | 29.90  | 29.95  | 31.42  | 31.52  | 31.76  | 31.40  | 32.86  | 33.22  | 33.35   |
| Average PSNR                    | 33.41  | 33.48  | 33.48  | 34.32  | 35.33  | 35.65  | 35.49  | 36.43  | 36.88  | 37.26   |
| Average SSIM                    | 0.8483 | 0.8511 | 0.8512 | 0.9113 | 0.9250 | 0.9280 | 0.9126 | 0.9364 | 0.9481 | 0.9505  |

- [784] Seonghyeon Nam, Youngbae Hwang, Yasuyuki Matsushita, and Seon Joo Kim. A holistic approach to cross-channel image noise modeling and its application to image denoising. *Proc. Computer Vision and Pattern Recognition (CVPR)*, pages 1683–1691, 2016. 1, 2, 3, 5, 6, 8
- [785] Glenn E Healey and Raghava Kondepudy. Radiometric ccd camera calibration and noise estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 16(3):267–276, 1994. 1, 3
- [786] J. Portilla. Full blind denoising through noise covariance estimation using gaussian scale mixtures in the wavelet domain. *Image Processing, 2004. ICIP '04. 2004 International Conference on*, 2:1217–1220, 2004. 1, 3
- [787] Tamer Rabie. Robust estimation approach for blind denoising. *Image Processing, IEEE Transactions on*, 14(11):1755–1765, 2005. 1, 3
- [788] C. Liu, R. Szeliski, S. Bing Kang, C. L. Zitnick, and W. T. Freeman. Automatic estimation and removal of noise from a single image. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(2):299–314, 2008. 1, 3
- [789] Zheng Gong, Zuowei Shen, and Kim-Chuan Toh. Image restoration with mixed or unknown noises. *Multiscale Modeling & Simulation*, 12(2):458–487, 2014. 1, 3
- [790] M. Lebrun, M. Colom, and J.-M. Morel. Multiscale image blind denoising. *Image Processing, IEEE Transactions on*, 24(10):3149–3161, 2015. 1, 2, 3, 5

- [791] Fengyuan Zhu, Guangyong Chen, and Pheng-Ann Heng. From noise modeling to blind image denoising. *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2016. 1, 3
- [792] C. M. Bishop. *Pattern recognition and machine learning*. New York: Springer, 2006. 1
- [793] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4):600–612, 2004. 1, 6
- [794] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian. Color image denoising via sparse 3d collaborative filtering with grouping constraint in luminance-chrominance space. *IEEE International Conference on Image Processing*, 1, 2007. 2
- [795] Neatlab ABSoft. Neat image. <https://neatvideo.com/home>. 2, 3, 5
- [796] G. Yu, G. Sapiro, and S. Mallat. Solving inverse problems with piecewise linear estimators: From Gaussian mixture models to structured sparsity. *IEEE Transactions on Image Processing*, 21(5):2481–2499, 2012. 2
- [797] Daniel Glasner, Shai Bagon, and Michal Irani. Super-resolution from a single image. *ICCV*, 2009.
- [798] Maria Zontak, Inbar Mosseri, and Michal Irani. Separating signal from noise using patch recurrence across scales. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1195–1202, 2013.

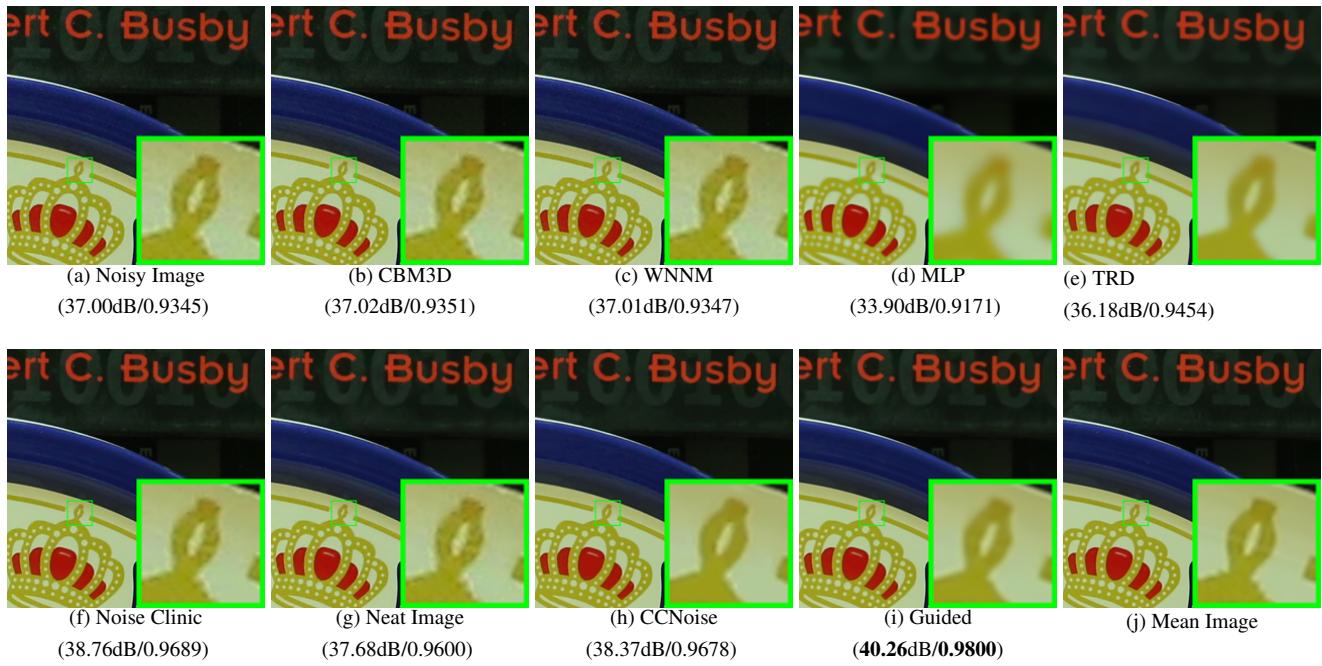


Figure 6. Denoised images of the image "Canon 5D Mark 3 ISO 3200 1" by different methods. The images are better to be zoomed in on screen.

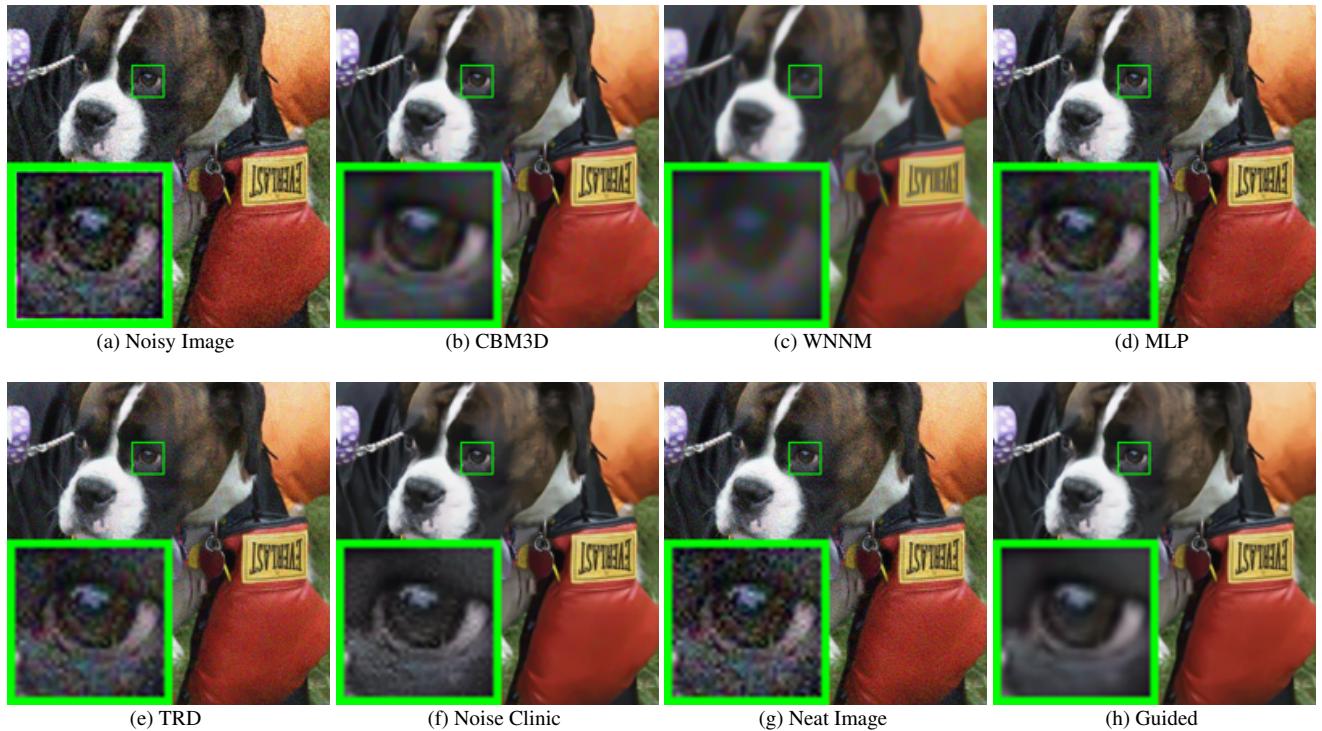


Figure 7. Denoised images of the image "5dmak3iso32003" by different methods. The images are better to be zoomed in on screen.

- [28] Tomer Michaeli and Michal Irani. Blind deblurring using internal patch recurrence. *European Conference on Computer Vision*, pages 783–798, 2014.
- [29] Y. Bahat and M. Irani. Blind dehazing using internal patch

recurrence. *IEEE International Conference on Computational Photography (ICCP)*, pages 1–9, May 2016.

- [30] J. Portilla, V. Strela, M.J. Wainwright, and E.P. Simoncelli. Image denoising using scale mixtures of Gaussians in the

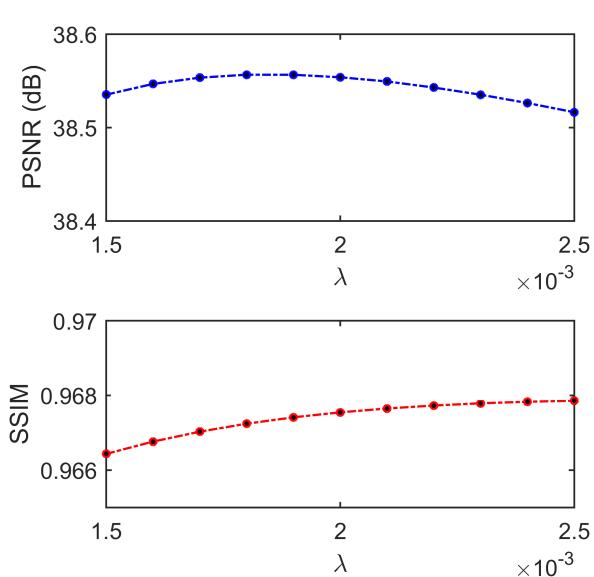


Figure 8. The PSNR/SSIM results as a function of the parameter  $\lambda$ .

wavelet domain. *Image Processing, IEEE Transactions on*, 12(11):1338–1351, 2003. 3

- [31] Peter J Huber. *Robust statistics*. Springer, 2011. 3
- [32] M. Lebrun, A. Buades, and J. M. Morel. A nonlocal bayesian image denoising algorithm. *SIAM Journal on Imaging Sciences*, 6(3):1665–1688, 2013. 3
- [33] W. Dong, L. Zhang, G. Shi, and X. Li. Nonlocally centralized sparse representation for image restoration. *Image Processing, IEEE Transactions on*, 22(4):1620–1630, 2013. 2, 3
- [34] Chenglong Bao, Jian-Feng Cai, and Hui Ji. Fast sparsity-based orthogonal dictionary learning for image restoration. *Proceedings of the IEEE International Conference on Computer Vision*, pages 3384–3391, 2013. 4
- [35] A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society. Series B (methodological)*, pages 1–38, 1977. 4
- [36] David L Donoho and Michael Elad. Optimally sparse representation in general (nonorthogonal) dictionaries via 1 minimization. *Proceedings of the National Academy of Sciences*, 100(5):2197–2202, 2003. 4
- [37] Franklin C. Crow. Summed-area tables for texture mapping. *SIGGRAPH Comput. Graph.*, 18(3):207–212, January 1984. 5
- [38] Paul Viola and Michael J Jones. Robust real-time face detection. *International journal of computer vision*, 57(2):137–154, 2004. 5
- [39] S. Osher, M. Burger, D. Goldfarb, J. Xu, and W. Yin. An iterative regularization method for total variation-based image restoration. *Multiscale Modeling & Simulation*, 4(2):460–489, 2005. 5