External Prior Guided Internal Prior Learning for Real Noisy Image Denoising

Anonymous CVPR submission

Paper ID 1047

Abstract

Existing image denoising methods mainly focus on signal independent noise such as Gaussian noise. The related paper showed that the internal priors (e.g., self-similarity) are highly effective for removing Gaussian noise from synthetic noisy images. However, the noise in real noisy images is much more complex than Gaussian noise since it is signal dependent and from several main sources. Hence, the internal priors learned on real noisy images would largely degrade the performance of denoising methods. In this paper, we propose to avoid the degradation by employing the external priors to guide the learning of internal priors. Specifically, we firstly learn external priors via Gaussian Mixture Model on the patch groups (PGs) extracted from natural clean images, and then employ the external priors to guide the subspace selection and orthogonal dictionary learning of the PGs from given noisy image. The orthogonal dictionary learning has closed-form solutions and the image denoising is simultaneously performed with the learning process. Extensive experiments on real noisy image denoising demonstrate that the proposed method achieves better performance than other state-of-the-art denoising methods.

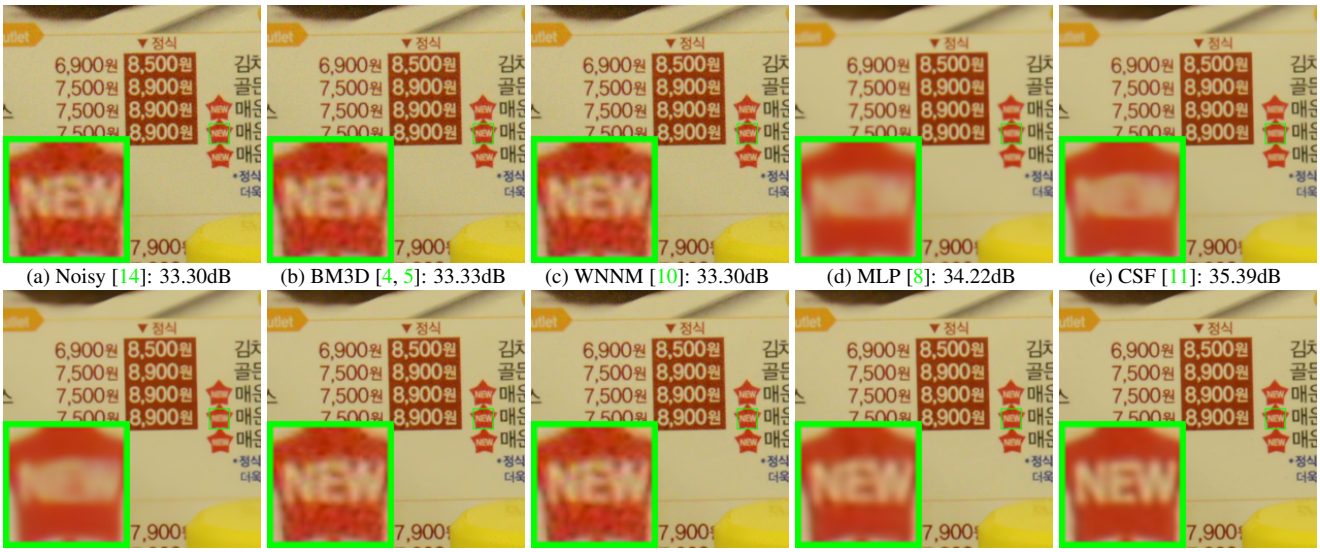
1. Introduction

Image denoising is an ideal platform for testing computer vision models and an essential procedure for the reliability of many vision systems such as medical imaging and surveillance. It aims to recover the latent clean image $\mathbf{x}$ from its noisy observation $\mathbf{y} = \mathbf{x}+\mathbf{n}$ where $\mathbf{n}$ is the additive noise. For several decades, there emerge numerous image denoising methods [1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13] most of which focusing on additive white Gaussian noise (AWGN). These methods exploited self-similarity [1, 7, 10, 12], structural sparsity [3, 4, 5, 6, 9], neural network [8], or learned filters [2, 11, 13] for image denoising problem.

However, the methods designed for synthetic Gaussian noise is much less effective for removing the noise in real noisy images. The reason is that the noise in real noisy images are much more complex than Gaussian [14, 15] and varies by different cameras and camera settings (ISO, shutter speed, and aperture, etc.). Besides, by the seminar work of [15], the noise in the camera imaging process [16, 14] are signal dependent and identified by five main sources: photon shot, fixed pattern, dark current, readout, and quantization noise. This would largely limit the performance of these methods, which are designed for signal independent noise, on dealing with complex real noise. To prove our argument, we apply some representing methods of BM3D [4], WNNM [10], MLP [8], CSF [11], and TRD [13] on a real noisy image provided in [14]. 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. The denoised images of these methods are listed in Figure 1 (from (b)$\sim$(f)). One can see that these methods either remove the noise or over-smooth the complex details in real noisy image.

In the last decade, there are also several methods [14, 17, 18, 19, 20, 21, 22, 23] developed for real noisy image denoising. Almost all these methods employ a two-stage framework: firstly estimating the parameters of the assumed noise model (usually Gaussian distribution or mixture of Gaussians (MoG)) and then performing denoising with the help of the estimated noise model. However, the noise in real noisy images are very complex and hard to be modeled by explict distributions. Therefore, the denoising performance of these methods would be influenced by the error between the estimated noise model and the actual noise distribution. This can be validated by the denoised results listed in Figure 1 ((g) and (h)), in which we evaluate the state-of-the-art real noisy image denoising method Noise Clinic [21, 22] and a commercial denoising software Neat Image [24]. From the results, we can see that the two methods do not perform well on the testing noisy image. These methods only exploit internal priors (e.g., noise prior, self-similarity) for real noisy image denoising and ignore the information of external natural clean images.

The external priors in natural clean 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



(a) Noisy [14]: 33.30dB (b) BM3D [4, 5]: 33.33dB (c) WNNM [10]: 33.30dB (d) MLP [8]: 34.22dB (e) CSF [11]: 35.39dB

(f) TRD [13]: 35.97dB (g) NC [21]: 35.33dB (h) NI [24]: 34.39dB (i) Ours: **37.02**dB (j) Mean Image [14]

Figure 1. Denoised images of the real noisy image "Nikon D800 ISO 3200 A3" from [14] by different methods. The images are better viewed by zooming in on screen.

priors of external clean images and priors of internalwould naturally achieves better performance. Based on above considerations, in this paper, we propose to employ the external priors learned from natural clean images to guide the learning of internal priors of given noisy image. The flowchart of the proposed method is illustrated in Fig. 2. In the external stage, we learn the external priors of patch groups [12] since nonlocal self-similarity is successfully used in many denoising methods [1, 4, 10, 12]. The learning of external priors is performed by Gaussian Mixture Model. In the internal stage, the learning of internal priors includes Gaussian selection and orthogonal dictionary learning (DL). The internal DL process includes two alternating stages: updating sparse coefficients and updating orthogonal dictionary. The image denoising is simultaneously done with the DL process. Both of the two stages have closed-form solutions. Hence, our proposed denoising method is very efficient. Through extensive experiments on real noisy image denoising, we demonstrate that the proposed method achieves better performance on real image denoising.

The rest of this paper will be summarized as follows: in Section 2, we briefly introduce the related work; in Section 3, we develop the proposed external prior guided internal prior learning model and formulate the overall image denoising algorithm; in Section 3, we perform extensive experiments on real image denoising problem; in Section 4, we conclude our paper and give future work.

2. Related Work

2.1. Internal *vs.* External Prior Learning

Image priors are playing key factors in image denoising methods [1, 3, 7, 12, 25]. There are mainly two categories of prior learning methods: 1) External methods prelearned priors (e.g., dictionaries) from a set of clean images, and the learned priors are used to recover the noisy images [7, 12]. 2) Internal methods directly learned priors from the given noisy image, and the image denoising is simultaneously done with the learning process [3, 26]. On one hand, the methods of [7, 12] demonstrated that the external priors learned from natural clean images are highly effctive and efficient for image denoising problem. On the other hand, the work of [9, 25] showed that the internal priors are highly effective for removing the signal independent Gaussian noise. However, the actual noise in real noisy images are much more complex than Gaussian noise. Only using the internal information may be not enough for real image denoising problem. The methods using only the external priors are not adaptive to given real noisy images, while the methods exploiting only internal priors would be degraded by the complex noise. In this paper, we propose to combine both the external and internal priors to solve the real noisy image denoising problem.

2.2. Real Image Denoising

In the last decade, researchers proposed many methods [17, 18, 19, 20, 21, 22, 23, 14] for real image denoising problem. In [17], Portilla proposed to use scale mixture of Gaussians to estimate the noise covariance matrix and latent clean images in wavelet domain. The work of Rabie [18] modeled the noisy pixels as outliers which are removed via Lorentzian robust estimator [27]. Liu *et al*. [19] proposed to use "noise level function" to estimate the noise and then use Gaussian conditional random field to obtain the latent clean image. Gong et al. [20] m odels the noise

by mixed ℓ_1 and ℓ_2 norms and remove the noise by sparsity prior in the wavelet transform domain. Later, Lebrun el al. proposed a multiscale denoising algorithm called "Noise Clinic" [21, 22]. This method generalizes the NL-Bayes model [28] to deal with blind noise and achieves state-of-the-art performance. Recently, Zhu et al. proposed a Bayesian model [23] which approximates and removes the noise via low-rank mixture of Gaussians. However, these methods largely depends on the modeling of noise in real noisy images which is hard to be modeled by explict distributions. Besides, the parametric estimation of the Gaussian or MoG distribution is often time consumming.

3. External Prior Guided Internal Prior Learning

In this section, we formulate the framework of external patch group (PG) prior guided internal prior learning. We first introduce the external PG prior leaning on natural clean RGB images. Then we propose to employ the learned external PG prior to guide the internal prior (subspace selection and dictionary learning (DL)) learning of given degraded (such as noisy) images. Under the weighted sparse coding framework, the internal prior learning process has alternative closed-form solutions in term of updating sparse coefficients and orthogonal dictionary. Finally, we discuss in details how external prior learned from natural clean images guide the internal prior learning of given degraded (noisy) images.

3.1. Learn External Patch Group Prior

In this section, we formulate the Patch Group prior learned on natural color images. Similar to [12], the patch group (PG) is defined as a group of similar patches to the local patch. The patch group mean is subtracted, and hence different groups of patches can share similar PGs. In this way, the space natural image patches to be modeled is largely reduced.

In this work, each local patch extracted from RGB images is of size $p \times p \times 3$. Then we search the M most similar patches $\{\mathbf{x}_m\}_{m=1}^M$ around each local patch through Euclidean distance, in a local window of size $W \times W$. The $\mathbf{x}_m \in \mathbb{R}^{3p^2 \times 1}$ is a patch vector formed by combining the 3 patch vectors (of size $p^2 \times 1$) in R, G, B channels. The mean vector of this PG is $\boldsymbol{\mu} = \frac{1}{M}\sum_{m=1}^M \mathbf{x}_m$, and the group mean subtracted PG is defined as $\overline{\mathbf{X}} \triangleq \{\overline{\mathbf{x}}_m = \mathbf{x}_m - \boldsymbol{\mu}\}, m = 1, ..., M$. Assume we have extracted N PGs from a set of external natural images, and the n-th PG is defined as $\overline{\mathbf{X}}_n \triangleq \{\overline{\mathbf{x}}_{n,m}\}_{m=1}^M, n = 1, ..., N$. We employ the Gaussian Mixture Model (GMM) to learn the external patch group based NSS prior. The overall objective log-likelihood function of GMM is

$$\ln \mathcal{L} = \sum_{n=1}^{N} \ln\Big(\sum_{k=1}^{K} \pi_k \prod_{m=1}^{M} \mathcal{N}(\overline{\mathbf{x}}_{n,m}|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)\Big). \quad (1)$$

The learning process is similar to the learning stage of the PGPD method [12]. Please refer to [12] for more details. After the learning stage, we finally obtain a GMM model with K Gaussian components. The learned parameters include mixture weights $\{\pi_k\}_{k=1}^K$, mean vectors $\{\boldsymbol{\mu}_k\}_{k=1}^K$, and covariance matrices $\{\boldsymbol{\Sigma}_k\}_{k=1}^K$. Similar to [12], the mean vector of each cluster is natural zeros, i.e., $\boldsymbol{\mu}_k = \mathbf{0}$.

Now, we have clustered the PGs extracted from external clean images into K Gaussians or subspaces. To better characterize each subspace, we perform singular value decomposition (SVD) on the covariance matrix:

$$\boldsymbol{\Sigma}_k = \mathbf{U}_k \mathbf{S}_k \mathbf{U}_k^\top. \quad (2)$$

The singular vector matrices $\{\mathbf{U}_k\}_{k=1}^K$ are employed as the external orthogonal dictionary to guide the internal dictionary learning. In Figure 3 (a) and (b), we illustrate an external clean image and one orthogonal dictionary learned via GMM on PGs of the external clean image. To better illustrate the dictionary The singular values in the diagonal of $\mathbf{S}_k$ reflect the significance of the singular vectors in $\mathbf{U}_k$ and utilized as prior weights for weighted sparse coding which will be discussed in next section.

3.2. External Prior Guided Internal Prior Learning

After the external patch group (PG) prior is learned, we can employ it to guide the internal PG prior learning for the given testing (real noisy) image. The guidance mainly comes from two aspects. One aspect is that the external prior can guide the internal noisy PGs to be assigned to most suitable Gaussians or subspaces. And for each subspace, the other aspect is to guide the orthogonal dictionary learning of internal noisy PGs.

3.2.1 Internal Subspace Selection

Given a real noisy image, assume we can totally extract N local patches from it. Similar to the external prior learning stage, for the n-th local patch ($n = 1, ..., N$), we extract its M most similar patches around it to form a noisy PG denoted by $\mathbf{Y}_n = \{\mathbf{y}_{n,1}, ..., \mathbf{y}_{n,M}\}$. Then the group mean of $\mathbf{Y}_n$, denoted by $\boldsymbol{\mu}_n$, is calculated and subtracted from each patch by $\overline{\mathbf{y}}_{n,m} = \mathbf{y}_{n,m} - \boldsymbol{\mu}_n$, leading to the mean subtracted PG $\overline{\mathbf{Y}}_n \triangleq \{\overline{\mathbf{y}}_{n,m}\}_{m=1}^M$. For adaptivity, we project the PG $\overline{\mathbf{Y}}_n$ into its most suitable Gaussian component (subspace) of the GMM learned on external PGs. The subspace most suitable for $\overline{\mathbf{Y}}_n$ is selected by firstly calculating the posterior probability of "$\overline{\mathbf{Y}}_n$ belonging to the k-th Gaussian component":

$$P(k|\overline{\mathbf{Y}}_n) = \frac{\prod_{m=1}^M \mathcal{N}(\overline{\mathbf{y}}_{n,m}|\mathbf{0}, \boldsymbol{\Sigma}_k)}{\sum_{l=1}^K \prod_{m=1}^M \mathcal{N}(\overline{\mathbf{y}}_{n,m}|\mathbf{0}, \boldsymbol{\Sigma}_l)} \quad (3)$$

for $k = 1, ..., K$, and then choosing the component with the maximum A-posteriori (MAP) probability $\max_k P(k|\overline{\mathbf{Y}}_n)$.

3.2.2 Internal Orthogonal Dictionary Learning

Assume we have assigned all internal noisy PGs $\{\overline{\mathbf{Y}}_n\}_{n=1}^N$ to their corresponding most suitable Gaussians or subspaces in $\{\mathcal{N}(\mathbf{0}, \mathbf{\Sigma}_k)\}_{k=1}^K$. For the k-th subspace, assume the noisy PGs assigned to it are $\{\overline{\mathbf{Y}}_{k_n}\}_{n=1}^{N_k}$ such that $\overline{\mathbf{Y}}_{k_n} = [\overline{\mathbf{y}}_{k_n,1}, ..., \overline{\mathbf{y}}_{k_n,M}]$ and $\sum_{k=1}^K N_k = N$. We utilize the external orthogonal dictionary $\mathbf{U}_k$ (Eqn. (2)) to guide the learning of an orthogonal dictionary for adaptively characterizing the internal PGs in the k-th subspace. The reasons we aim to learn orthogonal dictionaries are two-fold: firstly, given an orthonormal dictionary $\mathbf{D}$, its quality measure *mutual incoherence* $\mu(\mathbf{D}) = \max_{i=j} \frac{|\mathbf{d}_i^T \mathbf{d}_j|}{\|\mathbf{d}_i\|_2 \|\mathbf{d}_j\|_2}$ is naturally 0 and therefore better than other redundant dictionaries; secondly, the orthogonality of dictioanry can guarantee the dictionary learning stage has closed-form solutions (please refer to Eqn. (9)) and the closed-form solutions will make our proposed method efficient, which will be discussed in experimental section.

The learned dictionary $\mathbf{D}_k \triangleq [\mathbf{D}_{k,e}\ \mathbf{D}_{k,i}] \in \mathbb{R}^{3p^2 \times 3p^2}$ has two parts: the external part $\mathbf{D}_{k,e} = \mathbf{U}_k(:, 1 : 3p^2 - r) \in \mathbb{R}^{3p^2 \times (3p^2 - r)}$ is directly obtained from the external dictionary $\mathbf{U}_k$, and the internal part $\mathbf{D}_{k,i}$ is consisted of dictionary atoms adaptively learned from the internal noisy PGs $\{\overline{\mathbf{Y}}_{k_n}\}_{n=1}^{N_k}$. For notation simplicity, we ignore the subspace index k and denote the noisy PGs assigned to each subspace as $\mathbf{Y} \triangleq \{\overline{\mathbf{Y}}_n\}_{n=1}^N = [\overline{\mathbf{y}}_{1,1}, ..., \overline{\mathbf{y}}_{1,M}, ..., \overline{\mathbf{y}}_{N,1}, ..., \overline{\mathbf{y}}_{N,M}]$. The learning is performed under the weighted sparse coding framework proposed as follows:

$$\min_{\mathbf{D}_i, \{\boldsymbol{\alpha}_{n,m}\}} \sum_{n=1}^{N} \sum_{m=1}^{M} (\|\overline{\mathbf{y}}_{n,m} - \mathbf{D}\boldsymbol{\alpha}_{n,m}\|_2^2 + \sum_{j=1}^{3p^2} \lambda_j |\boldsymbol{\alpha}_{n,m,j}|)$$
$$s.t. \quad \mathbf{D} = [\mathbf{D}_e\ \mathbf{D}_i],\ \mathbf{D}_i^\top \mathbf{D}_i = \mathbf{I}_r,\ \mathbf{D}_e^\top \mathbf{D}_i = \mathbf{0}, \tag{4}$$

where $\boldsymbol{\alpha}_{n,m}$ is the sparse coefficient vector of the m-th patch $\overline{\mathbf{y}}_{n,m}$ in the n-th PG $\overline{\mathbf{Y}}_n$ and $\boldsymbol{\alpha}_{n,m,j}$ is the j-th element of $\boldsymbol{\alpha}_{n,m}$. λ_j is the j-th regularization parameter defined as

$$\lambda_j = \lambda / (\sqrt{\mathbf{S}_j} + \varepsilon). \tag{5}$$

We employ square roots of the singular values in $\mathbf{S}$ (please refer to Eqn. (2)) as external prior weights and add a small positive number ε to avoid zero denominator. Noted that $\mathbf{D}_e = \emptyset$ if $r = 3p^2$ and $\mathbf{D}_e = \mathbf{U}_k$ if $r = 0$. The dictionary $\mathbf{D} = [\mathbf{D}_e\ \mathbf{D}_i]$ is orthogonal by checking that:

$$\mathbf{D}^\top \mathbf{D} = \begin{bmatrix} \mathbf{D}_e^\top \\ \mathbf{D}_i^\top \end{bmatrix} [\mathbf{D}_e\ \mathbf{D}_i] = \begin{bmatrix} \mathbf{D}_e^\top \mathbf{D}_e & \mathbf{D}_e^\top \mathbf{D}_i \\ \mathbf{D}_i^\top \mathbf{D}_e & \mathbf{D}_i^\top \mathbf{D}_i \end{bmatrix} = \mathbf{I} \tag{6}$$

Similar to K-SVD [3], we employ an alternating iterative framework to solve the optimization problem (5). Specifically, we initialize the orthogonal dictionary as $\mathbf{D}_{(0)} = \mathbf{U}_k$ and for $t = 0, 1, ..., T-1$, alternatively do:

Updating Sparse Coefficient: given the orthogonal dictionary $\mathbf{D}_{(t)}$, we update the sparse coefficient vector of the m-th patch $\overline{\mathbf{y}}_{n,m}$ in the n-th PG $\overline{\mathbf{Y}}_n$ via solving

$$\boldsymbol{\alpha}_{n,m}^{(t)} := \arg\min_{\boldsymbol{\alpha}_{n,m}} \|\overline{\mathbf{y}}_{n,m} - \mathbf{D}_{(t)}\boldsymbol{\alpha}_{n,m}\|_2^2 + \sum_{j=1}^{3p^2} \lambda_j |\boldsymbol{\alpha}_{n,m,j}|$$
$$s.t. \quad \mathbf{D}_{(t)} = [\mathbf{D}_e\ \mathbf{D}_i],\ \mathbf{D}_i^\top \mathbf{D}_i = \mathbf{I}_r,\ \mathbf{D}_e^\top \mathbf{D}_i = \mathbf{0}, \tag{7}$$

Since dictionary $\mathbf{D}_{(t)} = [\mathbf{D}_e\ \mathbf{D}_i^{(t)}]$ is orthogonal, the problems (7) has a closed-form solution [12]

$$\boldsymbol{\alpha}_{n,m}^{(t)} = \text{sgn}(\mathbf{D}_{(t)}^\top \overline{\mathbf{y}}_{n,m}) \odot \max(|\mathbf{D}_{(t)}^\top \overline{\mathbf{y}}_{n,m}| - \mathbf{\Lambda}, \mathbf{0}), \tag{8}$$

where $\mathbf{\Lambda} = [\lambda_1, \lambda_2, ..., \lambda_{3p^2}]$ is the vector of regularization parameter and sgn(•) is the sign function, $\odot$ means element-wise multiplication.

Updating Internal Dictionary: given the sparse coefficient vectors $\mathbf{A}^{(t)} = [\boldsymbol{\alpha}_{1,1}^{(t)}, ..., \boldsymbol{\alpha}_{1,M}^{(t)}, ..., \boldsymbol{\alpha}_{N,1}^{(t)}, ..., \boldsymbol{\alpha}_{N,M}^{(t)}]$, we update the internal orthogonal dictionary via solving

$$\begin{aligned} \mathbf{D}_i^{(t+1)} &:= \arg\min_{\mathbf{D}_i} \sum_{n=1}^{N} \sum_{m=1}^{M} (\|\overline{\mathbf{y}}_{n,m} - \mathbf{D}\boldsymbol{\alpha}_{n,m}^{(t)}\|_2^2) \\ &= \arg\min_{\mathbf{D}_i} \|\mathbf{Y} - \mathbf{D}\mathbf{A}^{(t)}\|_F^2 \end{aligned} \tag{9}$$
$$s.t. \quad \mathbf{D} = [\mathbf{D}_e\ \mathbf{D}_i],\ \mathbf{D}_i^\top \mathbf{D}_i = \mathbf{I}_r,\ \mathbf{D}_e^\top \mathbf{D}_i = \mathbf{0},$$

The sparse coefficient matrix $\mathbf{A}^{(t)} = [(\mathbf{A}_e^{(t)})^\top\ (\mathbf{A}_i^{(t)})^\top]^\top$ also has two parts: the external part $\mathbf{A}_e^{(t)} \in \mathbb{R}^{(3p^2 - r) \times NM}$ and the internal part $\mathbf{A}_i^{(t)} \in \mathbb{R}^{r \times NM}$ denote the coefficients over external dictionary $\mathbf{D}_e$ and internal dictionary $\mathbf{D}_i^{(t)}$, respectively. According to the Theorem 4 in [30], the problem (9) has a closed-form solution $\mathbf{D}_i^{(t+1)} = \mathbf{U}_i \mathbf{V}_i^\top$, where $\mathbf{U}_i \in \mathbb{R}^{3p^2 \times r}$ and $\mathbf{V}_i \in \mathbb{R}^{r \times r}$ are the orthogonal matrices obtained by the following SVD

$$(\mathbf{I} - \mathbf{D}_e \mathbf{D}_e^\top) \mathbf{Y} (\mathbf{A}_i^{(t)})^\top = \mathbf{U}_i \mathbf{S}_i \mathbf{V}_i^\top. \tag{10}$$

The orthogonality of internal dictionary $\mathbf{D}_i^{(t+1)}$ can be checked by $(\mathbf{D}_i^{(t+1)})^\top (\mathbf{D}_i^{(t+1)}) = \mathbf{V}_i \mathbf{U}_i^\top \mathbf{U}_i \mathbf{V}_i^\top = \mathbf{I}_r$. In Figure 3 (c) and (d), we illustrate a denoised image by our proposed method and one internal orthogonal dictionary learned from PGs of the given noisy image.

3.3. The Denoising Algorithm

We evaluate the performance of the proposed framework on denoising real noisy images. The denoising is simultaneously done with the guided internal dictionary learning process. We ignore the index $k \in \{1, ..., K\}$ of subspace for notation simplicity. In the denoising stage, for each subspace, the group mean vectors $\{\boldsymbol{\mu}_n\}_{n=1}^N$ of corresponding mean subtracted noisy PGs $\{\overline{\mathbf{Y}}_n\}_{n=1}^N$ are saved for reconstruction. Until now, we obtain the solutions of sparse coefficient vectors $\{\hat{\boldsymbol{\alpha}}_{n,m}^{(T-1)}\}$ in Eqn. (8) for $n = 1, ..., N; m =$

Alg. 1: External Patch Group (PG) Prior Guided Internal PG Prior Learning for Image Denoising

Input: Noisy image $\mathbf{y}$, external PG prior GMM model
Output: The denoised image $\hat{\mathbf{x}}$.
Initialization: $\hat{\mathbf{x}}^{(0)} = \mathbf{y}$;
for $Ite = 1 : IteNum$ **do**
1. Extracting internal PGs from $\hat{\mathbf{x}}^{(Ite-1)}$;
   for each PG $\mathbf{Y}_n$ **do**
2. Calculate group mean vector $\boldsymbol{\mu}_n$ and form mean subtracted PG $\overline{\mathbf{Y}}_n$;
3. Subspace selection via Eqn. (3);
   end for
   for the PGs in each Subspace **do**
4. External PG prior Guided Internal Orthogonal Dictionary Learning by solving (4);
5. Recover each patch in all PGs via Eqn. (11);
   end for
6. Aggregate the recovered PGs of all subspaces to form the recovered image $\hat{\mathbf{x}}^{(Ite)}$;
end for

$1, ..., M$ and the orthogonal dictionary $\mathbf{D}_{(T)} = [\mathbf{D}_e \ \mathbf{D}_i^{(T)}]$ in Eqn. (9). Then the m-th latent clean patch $\hat{\mathbf{y}}_{n,m}$ in the n-th PG $\mathbf{Y}_n$ is recovered by

$$\hat{\mathbf{y}}_{n,m} = \mathbf{D}_{(T)}\hat{\boldsymbol{\alpha}}_{n,m} + \boldsymbol{\mu}_n, \tag{11}$$

where $n = 1, ..., N; m = 1, ..., M$. The latent clean image $\hat{\mathbf{x}}$ is reconstructed by aggregating all the estimated PGs. Similar to [12], we perform the above denoising procedures for several iterations for better denoising outputs. The proposed denoising algorithm is summarized in Alg. 1.

4. Experiments

In this section, we evaluate the performance of the proposed algorithm on real image denoising. To evaluation the effectiveness of the proposed framework of external prior guided internal prior learning, we compare it with the methods with only external prior or only internal prior (Section 4.3). We also compare the proposed algorithm with other state-of-the-art denoising methods [4, 5, 8, 10, 11, 13, 14, 21, 22, 24] (Section 4.4).

4.1. The Testing Datasets

The comparisons are performed on two standard datasets in which the images were captured under indoor or outdoor lighting conditions by different types of cameras and camera settings. The first dataset provided in [22] includes 20 real noisy images collected under uncontrolled outdoor environment. This dataset does not have "ground truth" images and hence the objective measurements can not be performed. In order to evaluate the compared methods on quantitative measures, we perform experiments on the second dataset provided in [14]. It includes 17 real noisy images and corresponding mean images. The noisy images were collected under controlled indoor environment. Some samples can be found in [14]. For each image, the same scene was shot 500 times under the same camera and camera setting. The mean image of the 500 shots is roughly taken as the "ground truth", with which the PSNR can be computed. Since the 17 images are too large (of size about $7000 \times 5000 \times 3$) and share repetitive contents, the authors in [14] performed comparison on 15 cropped images (of size $512 \times 521 \times 3$). To evaluate the compared methods on more samples, we cropped the 17 large images from [14] into 60 smaller images (of size $500 \times 500 \times 3$) including different contents. Some samples are shown in Figure 4. Note that the noise in our cropped 60 images used in [14] are different from the noise in the 15 images cropped by the authors of [14] since they are taken in different shots.

4.2. Implementation Details

Our proposed method contains two stages, the external prior learning stage and the external prior guided internal learning stage. In the first stage, we set $p = 6$ (so the patch size is $6 \times 6 \times 3$), $M = 10$ (the number of patches in a patch group (PG)), $W = 31$ (so the window size for PG searching is 31×31, and $K = 32$ (the number of Gaussians in Gaussian Mixture Model (GMM)). We learn the external prior via GMM on about 3.6 million PGs extracted from the Kodak PhotoCD Dataset (`http://r0k.us/graphics/kodak/`), which includes 24 high quality color images. In the second stage, we set $r = 54$ (the number of internal atoms in the learned dictionaries), $\lambda = 0.001$ (the sparse regularization parameter), $T = 2$ (the number of iterations for solving problem (4)), and $IteNum = 4$ (the number of iterations for Alg. 1). All experiments are performed under the Matlab2014b environment on a machine with Intel(R) Core(TM) i7-5930K CPU of 3.5GHz and 32GB RAM.

4.3. Comparison among external, internal and external guided internal priors

In this section, we compare our proposed method on real image denoising with external prior based method (denoted as "External") and internal prior based method (denoted as "Internal"). The three methods perform denoising under the same framework as Eqn. (5). The "External" method employs the external dictionaries (i.e., $r = 0$ in Eqn. (5)) for denoising, while the "Internal" method employ the GMM (also with $K = 32$ Gaussians) to directly cluster the internal noisy PGs into multiple subspaces, and for each subspace perform denoising with fully learned dictionary (i.e., $r = 3p^2$ in Eqn. (5)). All parameters of the three methods are tuned to achieve best performance.

We compare the above mentioned methods on the 60

(a) Noisy [14]: 35.89dB (b) External: 39.05dB (c) Internal: 38.75dB (d) Ours: **39.39**dB (e) Mean Image [14]

Figure 2. Denoised images of the 96-th cropped image from "Nikon D600 ISO 3200 C1" [14] by different methods. The images are better to be zoomed in on screen.

(a) Noisy [14]: 34.55dB (b) External: 36.09dB (c) Internal: 37.11dB (d) Ours: **37.39**dB (e) Mean Image [14]

Figure 3. Denoised images of the 94-th cropped image from "Nikon D600 ISO 3200 C1" [14] by different methods. The images are better to be zoomed in on screen.

Figure 4. Some samples cropped from real noisy images of [14].

Table 1. Average PSNR (dB) results and Run Time (seconds) of the External, the Internal, and our proposed methods on 60 real noisy images (of size $500 \times 500 \times 3$) cropped from [14].

	Noisy	**External**	**Internal**	**Ours**
PSNR	34.51	38.21	38.07	**38.75**
Time	—	**41.39**	667.36	44.06

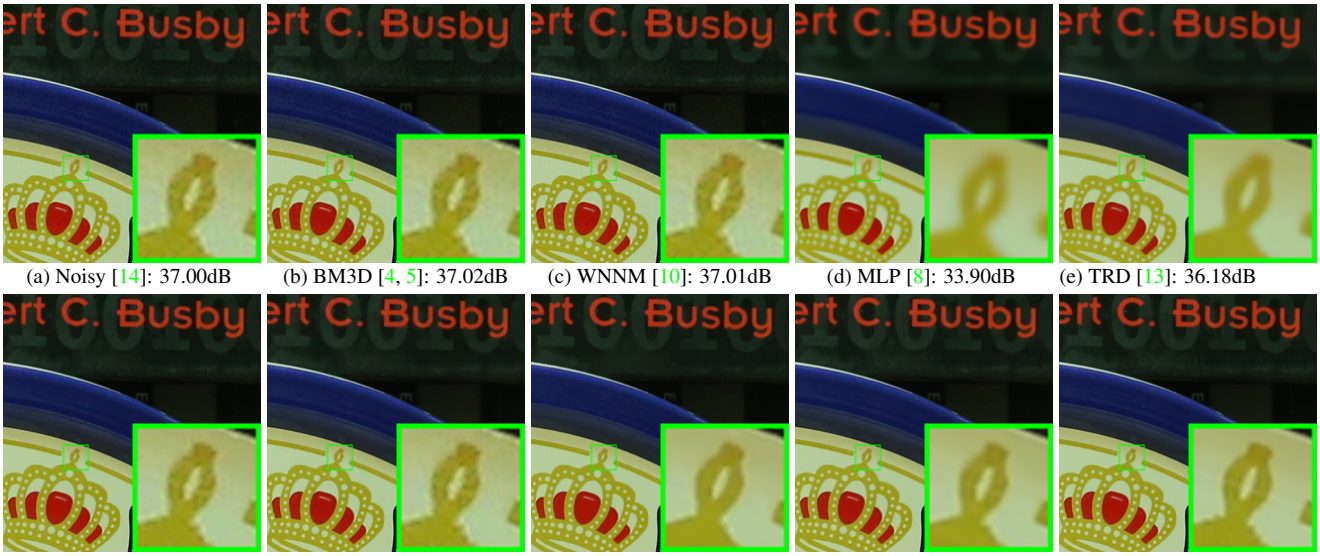
cropped images from [14]. The average PSNR and speed of these methods are listed in Table 1. It can be seen that our proposed method achieves better PSNR results than the methods of "External" and "Internal". The speed of our proposed method is much faster than the "Internal" method while only a little slower than the "External" method. We also compare the visual quality of the denoised images by these methods. From the results listed in Figure 2 and Figure 3, we can see that the "External" method is good at recovering structures (Figure 2) while the "Internal" method is good at recovering internal complex textures (Figure 3). And by utilizing both the external and internal priors, our proposed method can recover well both the structures and textures. Noted that the noisy images in Figures 2 and 3 are cropped from the same image captured by Nikon D600 at ISO $=$ 3200 in [14]. Hence, the differences on PSNR and visual quality among these methods only depends on the contents of the cropped images.

4.4. Comparison with Other Denoising Methods

In this section, we compare the proposed method with other state-of-the-art image denoising methods such as BM3D [4], WNNM [10], MLP [8], CSF [11], TRD [13], Noise Clinic (NC) [21], Cross-Channel (CC) [14], and Neat Image (NI) [24]. The methods of BM3D [4], WNNM [10], MLP [8], CSF [11], and TRD [13] are designed for removing Gaussian noise. For BM3D and WNNM, the level σ of Gaussian noise is very important and is estimated by the method [31]. The other parameters are set as default. For the methods of MLP, CSF, and TRD, we employ their default parameters settings. Since these methods are designed for grayscale images, we utilize them to denoise the R, G, B channels separately for color noisy images. The Noise Clinic (NC) [21] is a blind image denoising method which does not need any noise prior. We also compare with Neat Image (NI), a commercial software for image denoising. Due to its excellent performance, Neat Image (NI) is embedded into Photoshop and Corel PaintShop [24]. The comparisons are performed on the real noisy images from [22] and [14].

4.4.1 Comparison on the First Dataset [22]

The real noisy images in the first dataset [22] do not have "ground truth" images. We compare the proposed method with other state-of-the-art image denoising methods such as BM3D [4], WNNM [10], MLP [8], CSF [11], TRD [13],



(a) Noisy [14]: 37.00dB (b) BM3D [4, 5]: 37.02dB (c) WNNM [10]: 37.01dB (d) MLP [8]: 33.90dB (e) TRD [13]: 36.18dB

(f) NC [21]: 38.76dB (g) NI [24]: 37.68dB (h) CC [14]: 38.37dB (i) Ours: **40.50**dB (j) Mean Image [14]

Figure 5. 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.

Table 2. Average PSNR(dB) results of different methods on 60 real noisy images cropped from [14].

	Noisy	BM3D	WNNM	MLP	CSF	TRD	NI	NC	Ours
PSNR	34.51	34.58	34.52	36.19	37.40	37.75	36.53	37.57	38.75

Noise Clinic (NC) [21], Cross-Channel (CC) [14].

For the second dataset [22] with no "ground truth" images, we only compare the visual quality of the denoised images. Figure 6 shows the denoised images of "Dog" by the competing methods. More visual comparisons can be found in the supplementary file. It can be seen that the methods of BM3D, WNNM tend to globally over-smooth the image while locally remain some noise, while the methods of MLP, TRD are likely to remain noise in the whole image. This demonstrates that the methods designed for Gaussian noise are not effective for removing the complex noise in real noisy images. Though Noise Clinic and Neat Image are specifically developed for removing complex noise, they would sometimes fail to recover real noisy images. However, our proposed method recoveries more faithfully the structures and textures (such as the eye area) than the other competing methods.

4.4.2 Comparison on the Second Dataset [14]

while those in the second dataset [14] have corresponding "ground truth" images.

In this section, we compare the proposed method with other state-of-the-art image denoising methods such as BM3D [4], WNNM [10], MLP [8], CSF [11], TRD [13], Noise Clinic (NC) [21], Cross-Channel (CC) [14]. These methods are designed for removing Gaussian noise. For BM3D and WNNM, the level σ of Gaussian noise is very important and is estimated by the method [31]. The other parameters are set as default. For the methods of MLP, CSF, and TRD, we employ their default parameters settings. Since these methods are designed for grayscale images, we utilize them to denoise the R, G, B channels separately for color noisy images. The Noise Clinic (NC) [21] is a blind image denoising method which does not need any noise prior. The recently proposed Cross-Channel (CC) [14] is a state-of-the-art method on real image denoising problem. We also compare with Neat Image (NI), a commercial software for image denoising. Due to its excellent performance, NI is embedded into Photoshop and Corel PaintShop [24].

For the first dataset [14], the average PSNR results on the 60 cropped images are listed in Table 2 (the code of [14] is not available so that it is not compared). The numbers in red color and blue color are the best and second best results, respectively. It can be seen that our proposed method achieves much better PSNR results than the other methods. The improvement of our method over the second best method (TRD) is 1dB. To compare with [14], we also perform experiments on the 15 cropped images used in [14]. The PSNR values are listed in Table 3. As we can see, on most (9 out of the 15) images captured by different cameras and camera settings, our proposed method obtains better PSNR values than the other methods. Noted that, though in [14] a specific model is trained for each camera and camera setting, our proposed general method still gains 0.28dB improvements on PNSR over [14]. We also compare the visual quality of the denoised images by the competing methods. Figure 5 shows the denoised images of a scene captured by

Canon 5D Mark 3 at ISO $= 3200$ by the competing methods. More visual comparisons can be found in the supplementary file. We can see that BM3D, WNNM, NC, NI, and CC would either remain noise or generate artifacts, while MLP, TRD are likely to over-smooth the image. By combining the external and internal priors, our proposed method preserves edges and textures better than other methods.

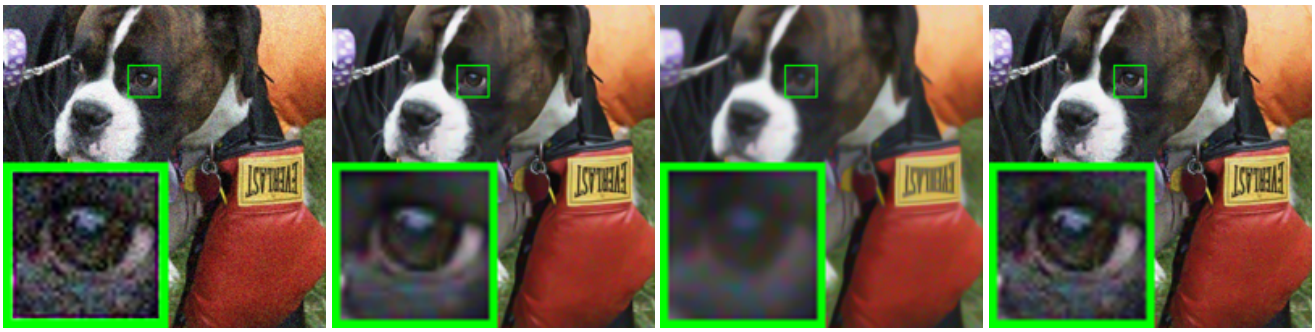
In summary, our proposed method demonstrates impressing ability on dealing with the complex noise in real noisy images in terms of PSNR and visual quality.

5. Conclusion and Future Work

Image priors are important for solving image denoising problems. The external priors learned from external clean images are generally effective to most images, while the internal priors learned directly from the noisy image are adaptive to the given image but would be biased by the complex noise in real noisy images. In this paper, we demonstrates that, once unifying both the priors in external clean images and internal noisy images, we can achieve much better while still efficient performance on real image denoising problem. Specifically, the external patch group (PG) priors learned on natural clean images can be used to guide the subspace selection and orthogonal dictionary learning of internal noisy PGs from given noisy images. The experiments on real image denoising problem have demonstrated the powerful ability of the proposed method. In the future, we will speed up the proposed algorithm and evaluate the proposed method on other computer vision tasks such as image super-resolution.

Table 3. Average PSNR(dB) results of different methods on 15 cropped real noisy images used in [14].

Camera Settings	**Noisy**	**BM3D**	**WNNM**	**MLP**	**CSF**	**TRD**	**NI**	**NC**	**CC**	**Ours**
Canon 5D Mark III 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	37.05
	33.83	33.88	33.90	32.37	32.63	33.10	34.77	35.54	34.91	36.11
Nikon D600 ISO = 3200	33.28	33.33	33.34	31.93	31.78	32.28	34.12	35.57	34.98	34.88
	33.77	33.85	33.79	34.15	35.16	35.34	35.36	36.70	35.95	36.31
	34.93	35.02	34.95	37.89	39.98	40.51	38.68	39.28	41.15	39.23
Nikon D800 ISO = 1600	35.47	35.54	35.57	33.77	34.84	35.09	37.34	38.01	37.99	38.40
	35.71	35.79	35.77	35.89	38.42	38.65	38.57	39.05	40.36	40.92
	34.81	34.92	34.95	34.25	35.79	35.85	37.87	38.20	38.30	38.97
Nikon D800 ISO = 3200	33.26	33.34	33.31	37.42	38.36	38.56	36.95	38.07	39.01	38.66
	32.89	32.95	32.96	34.88	35.53	35.76	35.09	35.72	36.75	37.07
	32.91	32.98	32.96	38.54	40.05	40.59	36.91	36.76	39.06	38.52
Nikon D800 ISO = 6400	29.63	29.66	29.71	33.59	34.08	34.25	31.28	33.49	34.61	33.76
	29.97	30.01	29.98	31.55	32.13	32.38	31.38	32.79	33.21	33.43
	29.87	29.90	29.95	31.42	31.52	31.76	31.40	32.86	33.22	33.58
Average	33.41	33.48	33.48	34.32	35.33	35.65	35.49	36.43	36.88	37.16



(a) Noisy [22] (b) BM3D [4, 5] (c) WNNM [10] (d) MLP [8]

(e) TRD [13] (f) NC [21] (g) NI [24] (h) Ours

Figure 6. Denoised images of the image "Dog" by different methods. The images are better to be zoomed in on screen.

References

[1] A. Buades, B. Coll, and J. M. Morel. A non-local algorithm for image denoising. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 60–65, 2005. 1, 2

[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. *IEEE Transactions on Image Processing*, 15(12):3736–3745, 2006. 1, 2, 4

[4] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian. Image denoising by sparse 3-D transform-domain collaborative filtering. *IEEE Transactions on Image Processing*, 16(8):2080–2095, 2007. 1, 2, 5, 6, 7, 9

[5] 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 (ICIP)*, pages 313–316, 2007. 1, 2, 5, 7, 9

[6] J. Mairal, F. Bach, J. Ponce, G. Sapiro, and A. Zisserman. Non-local sparse models for image restoration. *IEEE International Conference on Computer Vision (ICCV)*, pages 2272–2279, 2009. 1

[7] D. Zoran and Y. Weiss. From learning models of natural image patches to whole image restoration. *IEEE International Conference on Computer Vision (ICCV)*, pages 479–486, 2011. 1, 2

[8] H. C. Burger, C. J. Schuler, and S. Harmeling. Image denoising: Can plain neural networks compete with BM3D? *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2392–2399, 2012. 1, 2, 5, 6, 7, 9

[9] W. Dong, L. Zhang, G. Shi, and X. Li. Nonlocally centralized sparse representation for image restoration. *IEEE Transactions on Image Processing*, 22(4):1620–1630, 2013. 1, 2

[10] S. Gu, L. Zhang, W. Zuo, and X. Feng. Weighted nuclear norm minimization with application to image denoising. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2862–2869, 2014. 1, 2, 5, 6, 7, 9

[11] U. Schmidt and S. Roth. Shrinkage fields for effective image restoration. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2774–2781, June 2014. 1, 2, 5, 6, 7

[12] J. Xu, L. Zhang, W. Zuo, D. Zhang, and X. Feng. Patch group based nonlocal self-similarity prior learning for image denoising. *IEEE International Conference on Computer Vision (ICCV)*, pages 244–252, 2015. 1, 2, 3, 4, 5

[13] Y. Chen, W. Yu, and T. Pock. On learning optimized reaction diffusion processes for effective image restoration. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5261–5269, 2015. 1, 2, 5, 6, 7, 9

[14] S. Nam, Y. Hwang, Y. Matsushita, and S. J. Kim. A holistic approach to cross-channel image noise modeling and its application to image denoising. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1683–1691, 2016. 1, 2, 5, 6, 7, 9

[15] G. E Healey and R. Kondepudy. Radiometric CCD camera calibration and noise estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 16(3):267–276, 1994. 1

[16] 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, 2012. 1

[17] J. Portilla. Full blind denoising through noise covariance estimation using Gaussian scale mixtures in the wavelet domain. *IEEE International Conference on Image Processing (ICIP)*, 2:1217–1220, 2004. 1, 2

[18] T. Rabie. Robust estimation approach for blind denoising. *IEEE Transactions on Image Processing*, 14(11):1755–1765, 2005. 1, 2

[19] 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, 2

[20] Z. Gong, Z. Shen, and K.-C. Toh. Image restoration with mixed or unknown noises. *Multiscale Modeling & Simulation*, 12(2):458–487, 2014. 1, 2

[21] M. Lebrun, M. Colom, and J.-M. Morel. Multiscale image blind denoising. *IEEE Transactions on Image Processing*, 24(10):3149–3161, 2015. 1, 2, 3, 5, 6, 7, 9

[22] M. Lebrun, M. Colom, and J. M. Morel. The noise clinic: a blind image denoising algorithm. http://www.ipol.im/pub/art/2015/125/. Accessed 01 28, 2015. 1, 2, 3, 5, 6, 7, 9

[23] F. Zhu, G. Chen, and P.-A. Heng. From noise modeling to blind image denoising. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2016. 1, 2, 3

[24] Neatlab ABSoft. Neat Image. https://ni.neatvideo.com/home. 1, 2, 5, 6, 7, 9

[25] 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. 2

[26] 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

[27] P. J. Huber. *Robust statistics*. Springer, 2011. 2

[28] 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

[29] 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.

[30] H. Zou, T. Hastie, and R. Tibshirani. Sparse principal component analysis. *Journal of Computational and Graphical Statistics*, 15(2):265–286, 2006. 4

[31] X. Liu, M. Tanaka, and M. Okutomi. Single-image noise level estimation for blind denoising. *IEEE transactions on Image Processing*, 22(12):5226–5237, 2013. 6, 7