

000
001
002054
055
056003

External Prior Guided Internal Prior Learning for Real Noisy Image Denoising

057
058004
005
006
007
008
009
010
011059
060
061
062
063
064
065

012 Anonymous CVPR submission

066

013 Paper ID 1047

067

Abstract

Most of existing image denoising methods use some statistical models such as additive white Gaussian noise (AWGN) to model the noise, and learn image priors from either external data or the noisy image itself to remove noise. However, the noise in real-world noisy images is much more complex than AWGN, and it is hard to be modeled by simple analytical distributions. Therefore, many state-of-the-art denoising methods in literature become much less effective when applied to real noisy images. In this paper, we develop a robust denoiser for real noisy image denoising without explicit assumption on noise models. Specifically, we first learn external priors from a set of clean natural images, and then use the learned external priors to guide the learning of internal latent priors from the given noisy image. The proposed method is simple yet highly effective. Experiments on real noisy images demonstrate that it achieves much better denoising performance than state-of-the-art denoising methods, including those designed for real noisy images.

1. Introduction

Image denoising is a crucial and indispensable step to improve image quality in digital imaging systems. In particular, with the decrease of size of CMOS/CCD sensors, noise is more easily to be corrupted and hence denoising is becoming increasingly important for high resolution imaging. In literature of image denoising, the observed noisy image is usually modeled as $\mathbf{y} = \mathbf{x} + \mathbf{n}$, where \mathbf{x} is the latent clean image and \mathbf{n} is the corrupted noise. Numerous image denoising methods [1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13] have been proposed in the past decades, including sparse representation and dictionary learning based methods [1, 2, 3], nonlocal self-similarity based methods [4, 5, 6, 3, 7], low-rank based methods [8], neural network based methods [9], and discriminative learning based methods [10, 11].

Most of the existing denoising methods [1, 2, 4, 5, 6, 3, 7, 8, 9, 10, 11, 12, 13] mentioned above assume noise \mathbf{n} to be additive white Gaussian noise (AWGN). Unfortunately, this assumption is too ideal to be true for real-world noisy im-

ages, where the noise is much more complex than AWGN [14, 15] and varies by different cameras and camera settings (ISO, shutter speed, and aperture, etc.). According to [15], the noise corrupted in the imaging process [is signal dependent and comes from five main sources: photon shot, fixed pattern, dark current, readout, and quantization noise. As a result, many advanced denoising methods in literature becomes much less effective when applied to real-world noisy images. Fig. 1 shows an example, where we apply some representative and state-of-the-art denoising methods, including CBM3D [6], WNNM [8], MLP [9], CSF [10], and TRD [11], to a real noisy image (captured by a Nikon D800 camera with ISO is 3200) provided in [14]. One can see that these methods either remain the noise or over-smooth the image details on this real noisy image.

There have been a few methods [16, 17, 18, 14, 19, 20, 21] developed for real noisy image denoising. Almost all of these methods follow a two-stage framework: first estimate the parameters of the assumed noise model (usually Gaussian or mixture of Gaussians (MoG)), and then perform denoising with the estimated noise model. Again, the noise in real noisy images is very complex and hard to be modeled by explicit distributions such as Gaussian and MoG. Fig. 1 also shows the denoised results of two state-of-the-art real noisy image denoising methods, Noise Clinic [19, 20] and Neat Image [21]. One can see that these two methods do not perform well on this noisy image either.

This work aims to develop a robust solution for real noisy image denoising without explicitly assuming certain noise models. To achieve this goal, we propose to first learn image priors from external clean images, and then employ the learned external priors to guide the learning of internal latent priors from the given noisy image. The flowchart of the proposed method is illustrated in Fig. 3. We first extract millions of patch groups from a set of high quality natural images, with which a Gaussian Mixture Model (GMM) is learned as the external prior. The learned GMM prior model is used to cluster the patch groups extracted from the given noisy image, and then a hybrid orthogonal dictionary (HOD) is learned as the internal prior for image denoising. Our proposed denoising method is simple and ef-

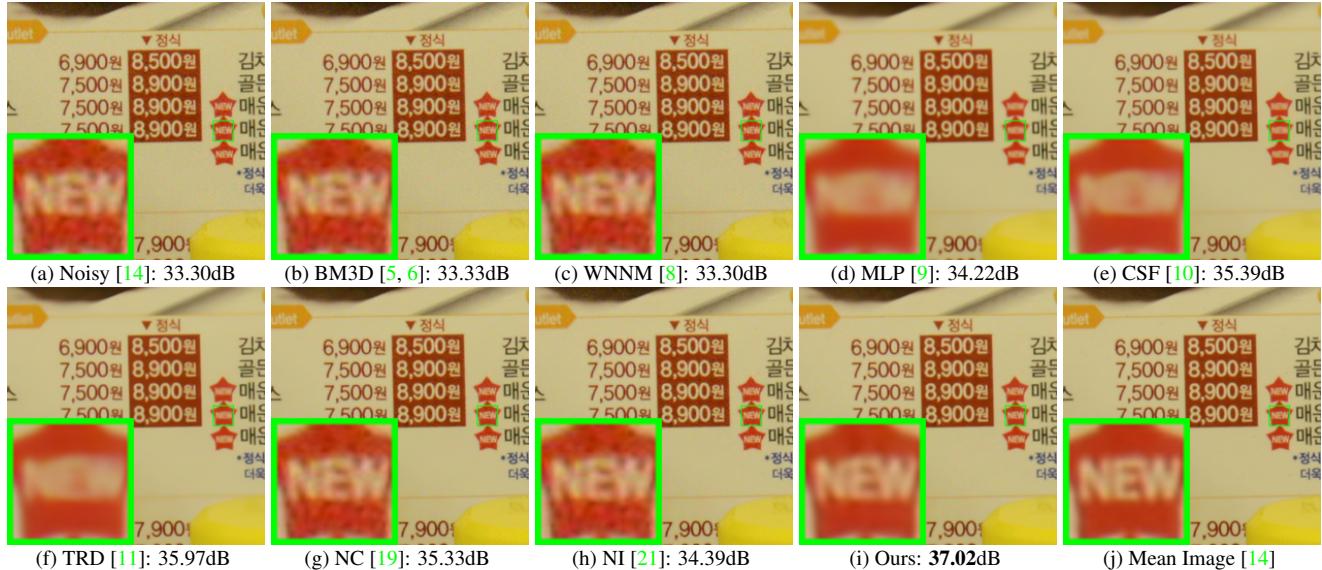


Figure 1. Denoised images of the real noisy image “Nikon D800 ISO 3200 A3” from [14] by different methods. The images are better viewed by zooming in on screen.

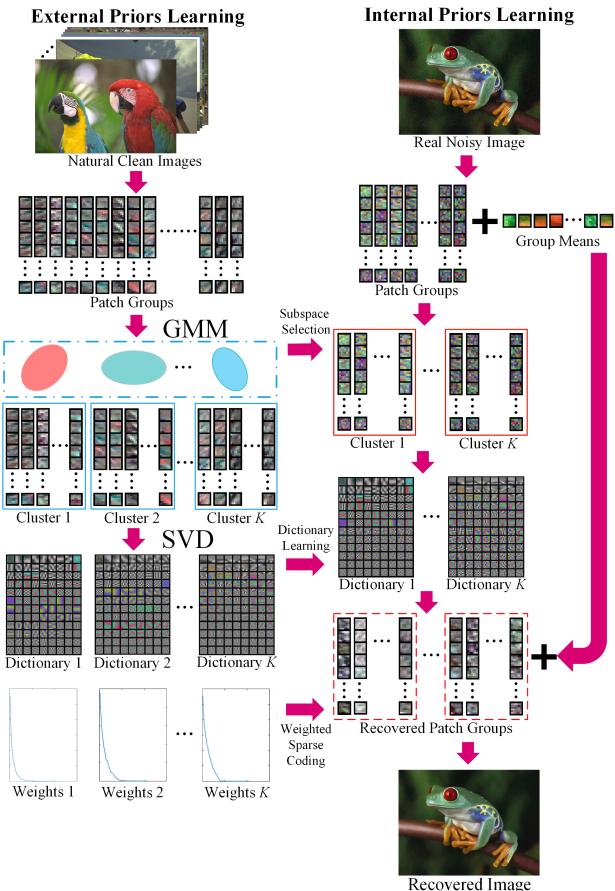


Figure 2. Flowchart of the proposed external prior guided internal prior learning and real noisy image denoising framework.

ficient, yet our extensive experiments on real noisy images clearly demonstrate its better denoising performance than the current state-of-the-arts.

2. Related Work

2.1. Internal vs. External Prior Learning

Image priors are playing a key role in image denoising [7, 13, 1, 22, 3, 23]. There are mainly two categories of prior learning methods. 1) External prior learning methods [12, 7, 13] learn priors (e.g., dictionaries) from a set of external clean images, and the learned priors are used to recover the latent clean image from noisy images. 2) Internal prior learning methods [1, 3, 22, 23] directly learn priors from the given noisy image, and the image denoising is often done simultaneously with the prior learning process. It has been demonstrated [7, 13] that the external priors learned from natural clean images are effective and efficient for image denoising problem, but they are not adaptive to the given noisy image so that some fine-scale image structures may not be well recovered. By contrast, the internal priors are adaptive to content of the given image, but the learning processing are usually slow. In addition, most of the internal prior learning methods [1, 3, 22, 23] assume AWGN noise, making the learned priors less robust for real noisy images. In this paper, we use external priors to guide the internal prior learning. Our method is not only much faster than the traditional internal learning methods, but also very effective to denoise real noisy images.

2.2. Real Noisy Image Denoising

In the last decade, there are many methods [16, 17, 19, 20, 18, 14] for blind image denoising problem. These methods can be applied to real noisy image denoising directly. Liu *et al.* [16] proposed to use “noise level function” to estimate the noise and then use Gaussian conditional random

216 field to obtain the latent clean image. Gong et al. [17] models the noise by mixed ℓ_1 and ℓ_2 norms and remove the
 217 noise by sparsity prior in the wavelet transform domain. Recently, Zhu et al. proposed a Bayesian model [18] which
 218 approximates and removes the noise via low-rank mixture
 219 of Gaussians. The method of “Noise Clinic” [19, 20] and the software of Neat Image [21] are developed specifically
 220 for real noisy image denoising. “Noise Clinic” [19, 20] generalizes the NL-Bayes model [24] to deal with blind noise
 221 and achieves state-of-the-art performance. However, these
 222 methods largely depends on the modeling of noise in real
 223 noisy images which is hard to be modeled by explicit distri-
 224 butions. Besides, the parametric estimation of the Gaussian
 225 or MoG distribution is often time consuming.
 226

231 3. External Prior Guided Internal Prior Learn- 232 ing

233 In this section, we first describe the learning of external
 234 prior, and then describe in detail the guided internal prior
 235 learning. Finally, the denoising algorithm with the learned
 236 priors is presented.

237 3.1. Learn External Patch Group Priors

238 The nonlocal self-similarity based patch group (PG) [7] has proved to be a very effective unit for image prior learning.
 239 In this work, we also extract PGs from natural clean images to learn priors. A PG is a group of similar patches to a local patch.

240 In our method, each local patch is extracted from a
 241 RGB image with patch size $p \times p \times 3$. We search the
 242 M most similar patches to this local patch (including the
 243 local patch itself) in a $W \times W$ local region around it.
 244 Each patch is stretched to a patch vector $\mathbf{x}_m \in \mathbb{R}^{3p^2 \times 1}$
 245 to form the PG $\{\mathbf{x}_m\}_{m=1}^M$. The mean vector of this PG is
 246 $\boldsymbol{\mu} = \frac{1}{M} \sum_{m=1}^M \mathbf{x}_m$, and the group mean subtracted PG is
 247 defined as $\bar{\mathbf{X}} \triangleq \{\bar{\mathbf{x}}_m = \mathbf{x}_m - \boldsymbol{\mu}\}$.

248 Assume we extract a number of N PGs from a set
 249 of external natural images, and the n -th PG is $\bar{\mathbf{Y}}_n \triangleq$
 250 $\{\bar{\mathbf{x}}_{n,m}\}_{m=1}^M, n = 1, \dots, N$. A Gaussian Mixture Model
 251 (GMM) is learned to model the PG prior. The overall log-
 252 likelihood function is

$$253 \ln \mathcal{L} = \sum_{n=1}^N \ln \left(\sum_{k=1}^K \pi_k \prod_{m=1}^M \mathcal{N}(\bar{\mathbf{x}}_{n,m} | \boldsymbol{\mu}_k, \Sigma_k) \right). \quad (1)$$

254 The learning process is similar to the GMM learning in
 255 [7, 13]. Finally, a GMM model with K Gaussian compo-
 256 nents is learned, and the learned parameters include mix-
 257 ture weights $\{\pi_k\}_{k=1}^K$, mean vectors $\{\boldsymbol{\mu}_k\}_{k=1}^K$, and covariance
 258 matrices $\{\Sigma_k\}_{k=1}^K$. Note that the mean vector of each
 259 cluster is naturally zero, i.e., $\boldsymbol{\mu}_k = \mathbf{0}$.

260 To better describe the subspace of each Gaussian com-
 261 ponent, we perform singular value decomposition (SVD)
 262 on the covariance matrix:

$$263 \Sigma_k = \mathbf{U}_k \mathbf{S}_k \mathbf{U}_k^\top. \quad (2)$$

264 The eigenvector matrices $\{\mathbf{U}_k\}_{k=1}^K$ will be employed as the
 265 external orthogonal dictionary to guide the internal dictio-
 266 nary learning in next sub-section. In Fig. 4 (a) and (b), we
 267 illustrate an external clean image and one orthogonal dictio-
 268 nary learned via GMM on PGs of the external clean im-
 269 age. The singular values in \mathbf{S}_k reflect the significance of
 270 the singular vectors in \mathbf{U}_k . They will also be utilized as
 271 prior weights for weighted sparse coding in our denoising
 272 algorithm.

273 3.2. Guided Internal Prior Learning

274 After the external PG prior is learned, we employ it to
 275 guide the internal PG prior learning for a given real noisy
 276 image. The guidance lies in two aspects. One is that the
 277 external prior can guide the subspace assignment of internal
 278 noisy PGs, while the other is that the external prior could
 279 guide the orthogonal dictionary learning of internal noisy
 280 PGs.

281 3.2.1 Internal Subspace Assignment

282 Given a real noisy image, we extract N (overlapped) lo-
 283 cal patches from it. Similar to the external prior learn-
 284 ing stage, for the n -th local patch we search its M most
 285 similar patches around it to form a noisy PG, denoted by
 286 $\mathbf{Y}_n = \{\mathbf{y}_{n,1}, \dots, \mathbf{y}_{n,M}\}$. Then the group mean of \mathbf{Y}_n ,
 287 denoted by $\boldsymbol{\mu}_n$, is subtracted from each patch by $\bar{\mathbf{y}}_{n,m} =$
 288 $\mathbf{y}_{n,m} - \boldsymbol{\mu}_n$, leading to the mean subtracted noisy PG $\bar{\mathbf{Y}}_n \triangleq$
 289 $\{\bar{\mathbf{y}}_{n,m}\}_{m=1}^M$.

290 The external GMM prior models $\{\Sigma_k\}_{k=1}^K$ basically
 291 characterize the subspaces of natural high quality PGs.
 292 Therefore, we project the noisy PG $\bar{\mathbf{Y}}_n$ into the subspaces
 293 of $\{\Sigma_k\}_{k=1}^K$ and assign it to the most suitable subspace
 294 based on the posterior probability:

$$295 P(k|\bar{\mathbf{Y}}_n) = \frac{\prod_{m=1}^M \mathcal{N}(\bar{\mathbf{y}}_{n,m} | \mathbf{0}, \Sigma_k)}{\sum_{l=1}^K \prod_{m=1}^M \mathcal{N}(\bar{\mathbf{y}}_{n,m} | \mathbf{0}, \Sigma_l)} \quad (3)$$

296 for $k = 1, \dots, K$. Then $\bar{\mathbf{Y}}_n$ is assigned to the component
 297 with the maximum A-posteriori (MAP) probability
 298 $\max_k P(k|\bar{\mathbf{Y}}_n)$.

299 3.2.2 Guided Orthogonal Dictionary Learning

300 Assume we have assigned all the internal noisy PGs
 301 $\{\bar{\mathbf{Y}}_n\}_{n=1}^N$ to their corresponding most suitable subspaces in
 302 $\{\mathcal{N}(\mathbf{0}, \Sigma_k)\}_{k=1}^K$. For the k -th subspace, the noisy PGs as-
 303 signed to it are $\{\bar{\mathbf{Y}}_{k,n}\}_{n=1}^{N_k}$ where $\bar{\mathbf{Y}}_{k,n} = [\bar{\mathbf{y}}_{k,n,1}, \dots, \bar{\mathbf{y}}_{k,n,M}]$
 304 and $\sum_{k=1}^K N_k = N$. We propose to learn an orthogonal dictio-
 305 nary \mathbf{D}_k from each set of PGs $\bar{\mathbf{Y}}_{k,n}$ with the guidance of
 306 the corresponding external orthogonal dictionary \mathbf{U}_k (Eqn.
 307 (2)) to characterize the internal PG prior. The reasons that
 308 we learn orthogonal dictionaries are two-fold. Firstly, the
 309 PGs $\bar{\mathbf{Y}}_{k,n}$ are in a subspace of the whole space of all PGs,
 310

324 therefore, there is no necessary to learn a redundant over-
 325 complete dictionary to characterize it, while an orthonormal
 326 dictionary has naturally zero *mutual incoherence* [25].
 327 Secondly, the orthogonality of dictionary can make the en-
 328 coding in the testing stage very efficient, leading to an ef-
 329 ficient denoising algorithm (please refer to sub-section 3.3
 330 for details).

331 We let the orthogonal dictionary \mathbf{D}_k be $\mathbf{D}_k \triangleq$
 332 $[\mathbf{D}_{k,E} \mathbf{D}_{k,I}] \in \mathbb{R}^{3p^2 \times 3p^2}$, where $\mathbf{D}_{k,E} = \mathbf{U}_k(:, 1 : 3p^2 -$
 333 $r) \in \mathbb{R}^{3p^2 \times r}$ is the external sub-dictionary and it includes
 334 the first r most important eigenvectors of \mathbf{U}_k , and the in-
 335 ternal sub-dictionary $\mathbf{D}_{k,I}$ is to be adaptively learned from
 336 the noisy PGs $\{\bar{\mathbf{Y}}_{k_n}\}_{n=1}^{N_k}$. The rationale to design \mathbf{D}_k as a
 337 hybrid dictionary is as follows. The external sub-dictionary
 338 $\mathbf{D}_{k,E}$ is pre-trained from external clean data, and it repre-
 339 sents the k -th latent subspace of natural images, which is
 340 helpful to reconstruct the common latent structures of im-
 341 ages. However, $\mathbf{D}_{k,E}$ is general to all images and it is not
 342 adaptive to the given noisy image. Some fine-scale details
 343 specific to the given image may not be well characterized by
 344 $\mathbf{D}_{k,E}$. Therefore, we learn an internal sub-dictionary $\mathbf{D}_{k,I}$ to
 345 supplement $\mathbf{D}_{k,E}$. In other words, $\mathbf{D}_{k,I}$ is to reveal the
 346 latent subspace adaptive to the input noisy image, which
 347 cannot be effectively represented by $\mathbf{D}_{k,E}$.

348 For notation simplicity, in the following development we
 349 ignore the subspace index k for $\bar{\mathbf{Y}}_{k_n}$ and \mathbf{D}_k , etc. The
 350 learning of hybrid orthogonal dictionary \mathbf{D} is performed un-
 351 der the following weighted sparse coding framework:

$$\begin{aligned} & \min_{\mathbf{D}_i, \{\alpha_{n,m}\}} \sum_{n=1}^N \sum_{m=1}^M (\|\bar{\mathbf{y}}_{n,m} - \mathbf{D}\alpha_{n,m}\|_2^2 + \sum_{j=1}^{3p^2} \lambda_j |\alpha_{n,m,j}|) \\ & \text{s.t. } \mathbf{D} = [\mathbf{D}_e \mathbf{D}_i], \mathbf{D}_i^\top \mathbf{D}_i = \mathbf{I}_r, \mathbf{D}_e^\top \mathbf{D}_i = \mathbf{0}, \end{aligned} \quad (4)$$

352 where $\alpha_{n,m}$ is the sparse coding vector of the m -th patch
 353 $\bar{\mathbf{y}}_{n,m}$ in the n -th PG $\bar{\mathbf{Y}}_n$ and $\alpha_{n,m,j}$ is the j -th element of
 354 $\alpha_{n,m}$. λ_j is the j -th regularization parameter defined as

$$\lambda_j = \lambda / (\sqrt{\mathbf{S}_k(j)} + \varepsilon), \quad (5)$$

355 where $\mathbf{S}_k(j)$ is the j -th singular value of diagonal singu-
 356 lar value matrix \mathbf{S}_k (please refer to Eqn. (2)) and ε is a
 357 small positive number to avoid zero denominator. Noted
 358 that $\mathbf{D}_E = \mathbf{U}_k$ if $r = 3p^2$ and $\mathbf{D}_E = \emptyset$ if $r = 0$. The
 359 dictionary $\mathbf{D} = [\mathbf{D}_E \mathbf{D}_I]$ is orthogonal by checking that:

$$\mathbf{D}^\top \mathbf{D} = \begin{bmatrix} \mathbf{D}_E^\top \\ \mathbf{D}_I^\top \end{bmatrix} [\mathbf{D}_E \mathbf{D}_I] = \begin{bmatrix} \mathbf{D}_E^\top \mathbf{D}_E & \mathbf{D}_E^\top \mathbf{D}_I \\ \mathbf{D}_I^\top \mathbf{D}_E & \mathbf{D}_I^\top \mathbf{D}_I \end{bmatrix} = \mathbf{I} \quad (6)$$

360 We employ an alternating iterative approach to solve the
 361 optimization problem (4). Specifically, we initialize the or-
 362 thogonal dictionary as $\mathbf{D}^{(0)} = \mathbf{U}_k$ and for $t = 0, 1, \dots, T -$
 363 1, we alternatively update $\alpha_{n,m}$ and \mathbf{D} as follows:

364 **Updating Sparse Coefficient:** Given the orthogonal dic-
 365 tionary $\mathbf{D}^{(t)}$, we update each sparse coding vector $\alpha_{n,m}$ by
 366 solving

$$\alpha_{n,m}^{(t)} := \arg \min_{\alpha_{n,m}} \|\bar{\mathbf{y}}_{n,m} - \mathbf{D}^{(t)} \alpha_{n,m}\|_2^2 + \sum_{j=1}^{3p^2} \lambda_j |\alpha_{n,m,j}| \quad (7)$$

367 Since dictionary $\mathbf{D}^{(t)}$ is orthogonal, the problems (7) has a
 368 closed-form solution

$$\alpha_{n,m}^{(t)} = \text{sgn}((\mathbf{D}^{(t)})^\top \bar{\mathbf{y}}_{n,m}) \odot \max(|(\mathbf{D}^{(t)})^\top \bar{\mathbf{y}}_{n,m}| - \lambda, 0), \quad (8)$$

369 where $\lambda = [\lambda_1, \lambda_2, \dots, \lambda_{3p^2}]$ is the vector of regulariza-
 370 tion parameter and $\text{sgn}(\bullet)$ is the sign function, \odot means
 371 element-wise multiplication. The detailed derivation of
 372 Eqn. (8) can be found in the supplementary file.

373 **Updating Internal Sub-dictionary:** Given the sparse cod-
 374 ing vectors $\alpha_{n,m}^{(t)}$, we update the internal sub-dictionary by
 375 solving

$$\begin{aligned} \mathbf{D}_i^{(t+1)} &:= \arg \min_{\mathbf{D}_i} \sum_{n=1}^N \sum_{m=1}^M (\|\bar{\mathbf{y}}_{n,m} - \mathbf{D}_i \alpha_{n,m}^{(t)}\|_2^2) \\ &= \arg \min_{\mathbf{D}_i} \|\mathbf{Y} - \mathbf{D} \mathbf{A}^{(t)}\|_F^2 \end{aligned} \quad (9)$$

$$\text{s.t. } \mathbf{D} = [\mathbf{D}_e \mathbf{D}_i], \mathbf{D}_i^\top \mathbf{D}_i = \mathbf{I}_r, \mathbf{D}_e^\top \mathbf{D}_i = \mathbf{0}, \quad (39)$$

400 The sparse coefficient matrix $\mathbf{A}^{(t)} = [(\mathbf{A}_e^{(t)})^\top \ (\mathbf{A}_i^{(t)})^\top]^\top$
 401 also has two parts: the external part $\mathbf{A}_e^{(t)} \in \mathbb{R}^{(3p^2-r) \times NM}$
 402 and the internal part $\mathbf{A}_i^{(t)} \in \mathbb{R}^{r \times NM}$ denote the coefficients
 403 over external dictionary \mathbf{D}_e and internal dictionary $\mathbf{D}_i^{(t)}$,
 404 respectively. According to the Theorem 4 in [26], the prob-
 405 lem (9) has a closed-form solution $\mathbf{D}_i^{(t+1)} = \mathbf{U}_i \mathbf{V}_i^\top$, where
 406 $\mathbf{U}_i \in \mathbb{R}^{3p^2 \times r}$ and $\mathbf{V}_i \in \mathbb{R}^{r \times r}$ are the orthogonal matrices
 407 obtained by the following SVD

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

411 The orthogonality of internal dictionary $\mathbf{D}_i^{(t+1)}$ can be
 412 checked by $(\mathbf{D}_i^{(t+1)})^\top (\mathbf{D}_i^{(t+1)}) = \mathbf{V}_i \mathbf{U}_i^\top \mathbf{U}_i \mathbf{V}_i^\top = \mathbf{I}_r$.
 413 In Figure 4 (c) and (d), we illustrate a denoised image by
 414 our proposed method and one internal orthogonal dictionary
 415 learned from PGs of the given noisy image.

3.3. The Denoising Algorithm

416 We evaluate the performance of the proposed framework
 417 on denoising real noisy images. The denoising is simultane-
 418 ously done with the guided internal dictionary learning
 419 process. We ignore the index $k \in \{1, \dots, K\}$ of subspace
 420 for notation simplicity. In the denoising stage, for each sub-
 421 space, the group mean vectors $\{\mu_n\}_{n=1}^N$ of corresponding
 422 mean subtracted noisy PGs $\{\bar{\mathbf{Y}}_n\}_{n=1}^N$ are saved for recon-
 423 struction. Until now, we obtain the solutions of sparse coef-
 424 ficient vectors $\{\hat{\alpha}_{n,m}^{(T-1)}\}$ in Eqn. (8) for $n = 1, \dots, N; m =$
 425 1, ..., M and the orthogonal dictionary $\mathbf{D}_{(T)} = [\mathbf{D}_e \mathbf{D}_i^{(T)}]$
 426 in Eqn. (9). Then the m -th latent clean patch $\hat{\mathbf{y}}_{n,m}$ in the
 427 n -th PG \mathbf{Y}_n is recovered by

$$\hat{\mathbf{y}}_{n,m} = \mathbf{D}_{(T)} \hat{\alpha}_{n,m} + \mu_n, \quad (11)$$

Alg. 1: External Prior Guided Internal Prior Learning
for Real Noisy 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 μ_n and form
 mean subtracted PG $\bar{\mathbf{Y}}_n$;
 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

where $n = 1, \dots, N; m = 1, \dots, M$. The latent clean image $\hat{\mathbf{x}}$ is reconstructed by aggregating all the estimated PGs. Similar to [7], 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 [5, 6, 9, 8, 10, 11, 14, 19, 20, 21] (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 [20] 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 cam-

era 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 $512 \times 512 \times 3$). 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 5. 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”). For the “External” method, we utilize the external dictionaries (i.e., $r = 0$ in Eqn. (5)) for denoising. For the given noisy image, we extract the PGs and then do internal subspace selection via Eqn. 3. The denoising is performed via the weighted sparse coding framework proposed in [7]. For the “Internal” method, the overall framework is similar to the method of [3]. We employ the GMM model (also with $K = 32$ Gaussians) to cluster the noisy PGs extracted from given noisy image into multiple subspaces, and for each subspace, we utilize the internal orthogonal dictionary obtained via Eqn. (2) by weighted sparse coding framework in [7]. All parameters of the three methods are tuned to achieve best performance.

We compare the above mentioned methods on the 60 cropped images (of size $500 \times 500 \times 3$) from [14]. The

486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539

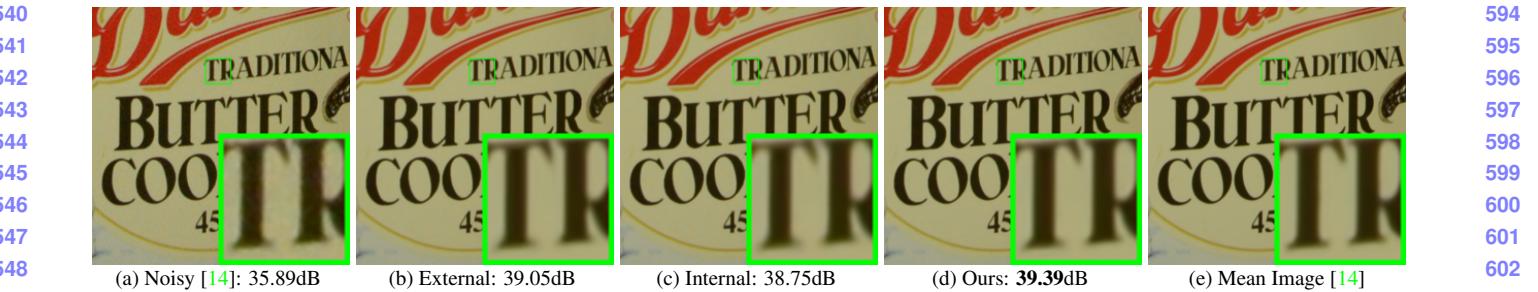


Figure 3. Denoised images of the 96-th cropped image from “Nikon D600 ISO 3200 C1” [14] by different methods. The images are better to be zoomed in on screen.

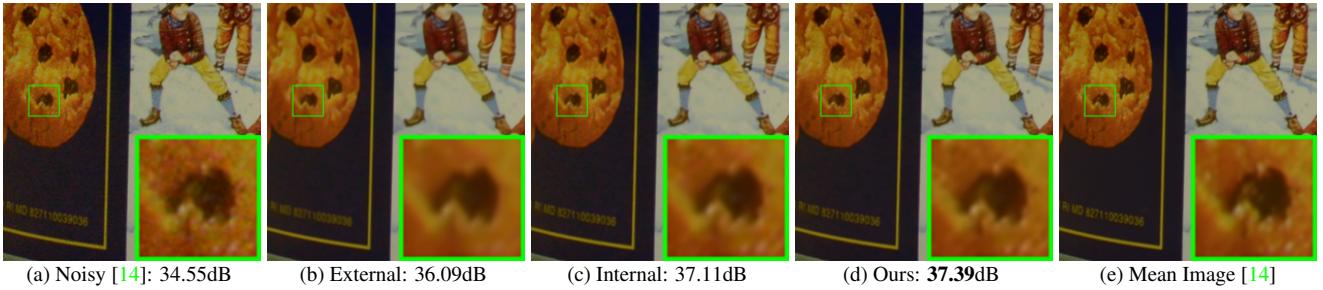


Figure 4. Denoised images of the 94-th cropped image from “Nikon D600 ISO 3200 C1” [14] by different methods. The images are better to be zoomed in on screen.



Figure 5. 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	—	39.57	667.36	41.89

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 3 and Figure 4, we can see that the “External” method is good at recovering structures (Figure 3) while the “Internal” method is good at recovering internal complex textures (Figure 4). 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 3 and 4 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 [5], WNNM [8], MLP [9], CSF [10], TRD [11], Noise Clinic (NC) [19], Cross-Channel (CC) [14], and Neat Image (NI) [21]. The methods of BM3D [5], WNNM [8], MLP [9], CSF [10], and TRD [11] 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 [27]. 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) [19] 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 [21]. The comparisons are performed on the real noisy images from [20] and [14].

4.4.1 Comparison on the First Dataset [20]

The real noisy images in the dataset [20] do not have “ground truth” images. On this dataset, we compare the proposed method with the methods of BM3D [5], WNNM [8], MLP [9], TRD [11], Noise Clinic (NC) [19], and Neat Image (NI) [21]. We only compare the visual quality of the denoised images. Figure 6 shows the denoised images

594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647

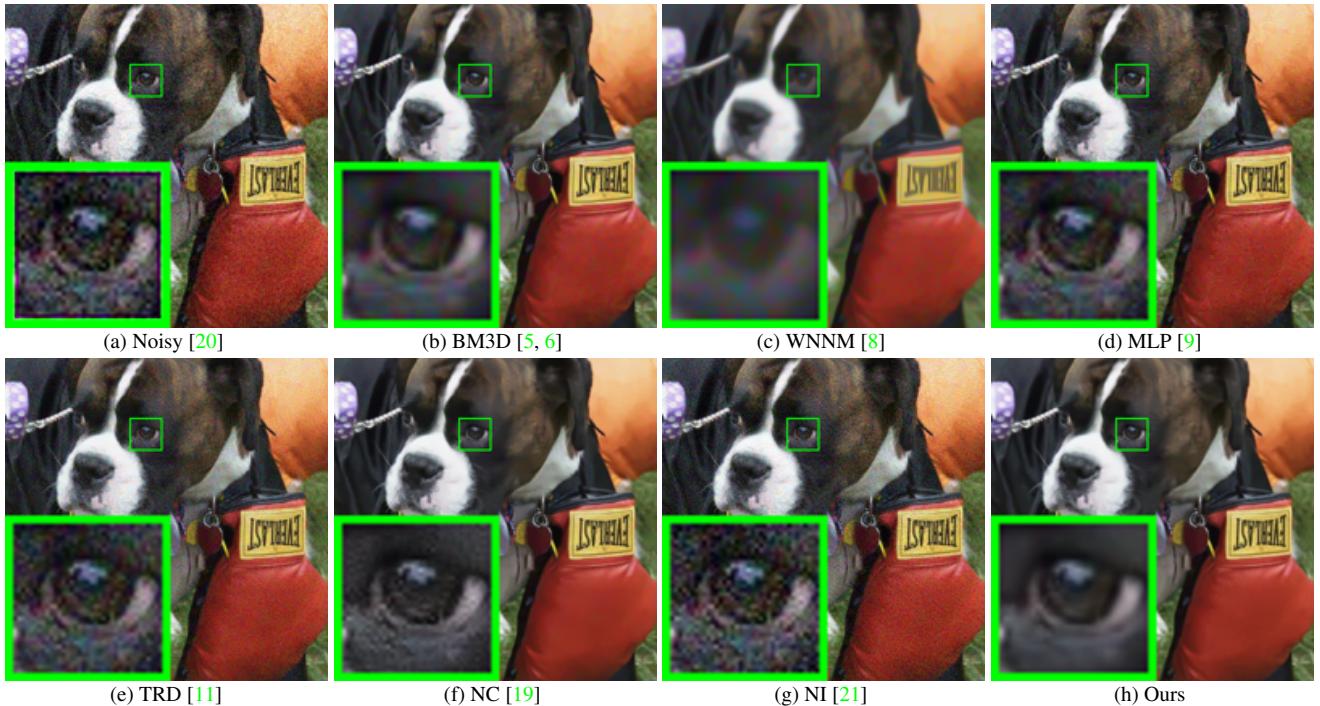


Figure 6. Denoised images of the image “Dog” by different methods. The images are better to be zoomed in on screen.

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]

The real noisy images in the second dataset [14] have corresponding “ground truth” images. On this dataset, we firstly perform comparison on the 15 cropped images used in [14]. The compared method are BM3D [5], WNNM [8], MLP [9], CSF [10], TRD [11], Noise Clinic (NC) [19], and Cross-Channel (CC) [14]. The PSNR values are listed in Table 2. 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 de-

noised images by the competing methods. Figure 7 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.

To evaluate the compared methods on more samples, we then perform denoising experiments on the 60 smaller images cropped from the 17 images provided in [14]. The average PSNR results are listed in Table 3 (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. Due to the spacial limitations, the visual comparisions are provided in the supplementary file.

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 demon-

Table 2. 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

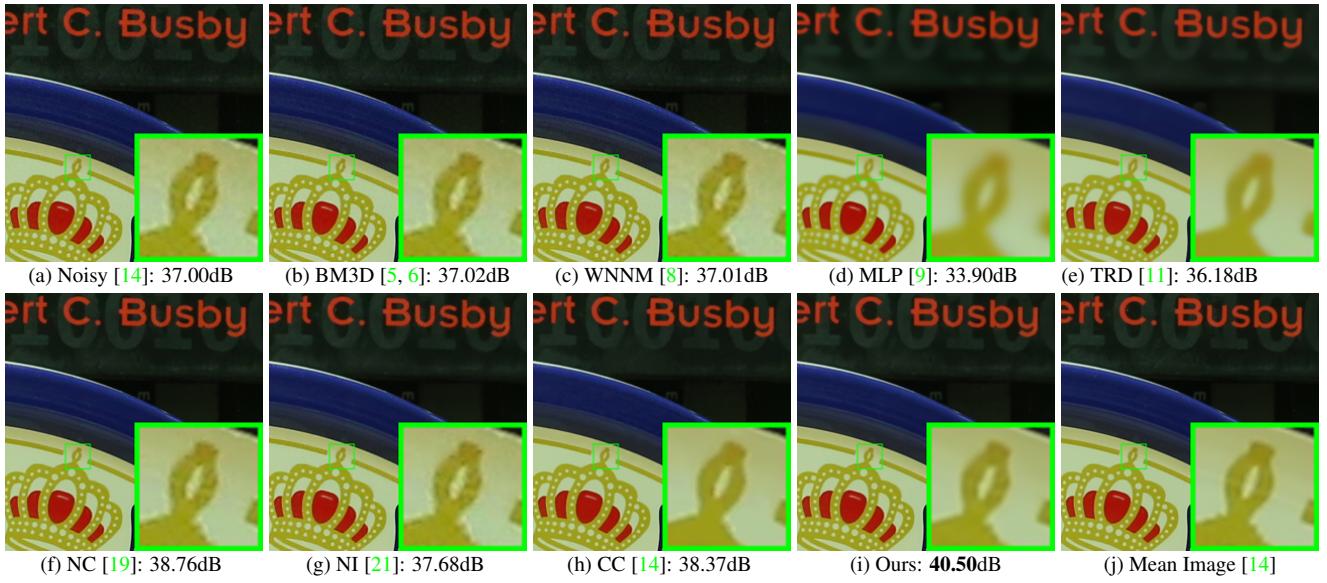


Figure 7. Denoised images of the image “Canon 5D Mark 3 ISO 3200” by different methods. The images are better to be zoomed in on screen.

Table 3. Average PSNR(dB) results of different methods on 60 real noisy images cropped from [14].

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

strates 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.

864
865
866
867
868
869**References**870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917

- [1] 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
- [2] 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
- [3] 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, 5
- [4] 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
- [5] 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, 8
- [6] 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, 8
- [7] 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, 5
- [8] 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, 8
- [9] 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, 8
- [10] 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
- [11] 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, 8
- [12] S. Roth and M. J. Black. Fields of experts. *International Journal of Computer Vision*, 82(2):205–229, 2009. 1, 2
- [13] 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, 3
- [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, 8
- [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] 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
- [17] 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, 3
- [18] 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
- [19] 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, 8
- [20] 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
- [21] Neatlab ABSoft. Neat Image. <https://ni.neatvideo.com/home>. 1, 2, 3, 5, 6, 7, 8
- [22] 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
- [23] M. Zontak and M. Irani. Internal statistics of a single natural image. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 977–984, 2011. 2
- [24] 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
- [25] D. L. Donoho and X. Huo. Uncertainty principles and ideal atomic decomposition. *IEEE Transactions on Information Theory*, 47(7):2845–2862, 2001. 4
- [26] H. Zou, T. Hastie, and R. Tibshirani. Sparse principal component analysis. *Journal of Computational and Graphical Statistics*, 15(2):265–286, 2006. 4
- [27] 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

918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971