

000
001
002
003
004
005
006
007
008
009
010
011054
055
056
057
058
059
060
061
062
063
064
065
066
067
068
069
070
071
072
073
074
075
076
077
078
079
080
081
082
083
084
085
086
087
088
089
090
091
092
093
094
095
096
097
098
099
100
101
102
103
104
105
106
107

A Noise Estimation Free Framework for Robust Real Image Denoising

Anonymous CVPR submission

Paper ID ****

Abstract

Existing image denoising methods largely depends on noise modeling and estimation. The commonly used noise models, additive white Gaussian or Mixture of Gaussians, are inflexible in describing the complex noise on real-world noisy images or time consuming in parametric estimation, respectively. Therefore, how to perform image denoising **without** noise modeling and estimation is an essential while challenging problem. In this paper, we attempt to solve this problem by directly learning the transformation process between the noisy images and clean ones. The transformation is learned on patches instead of images for dimensional tractability. The learning data is collected by constructing paired noisy and clean patches from unpaired real-world noisy and clean images. Since real noise is signal dependent and from several main sources [1], we cluster the learning data into multiple components. For each component, we learn in an integrated way two paired dictionaries for the noisy and clean data and two transformation functions between them. The overall learned transformation process could remove the noise from different sources. Experiments show that the proposed Paired Dictionary and Transformation Learing (PDTL) model achieves better performance on denoising real-world noisy images than existing noise estimation based methods.

1. Introduction

Image denoising is a fundamental problem in computer vision and image processing. It is 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. For several decades, there emerge numerous image denoising methods and most of them focus on dealing with additive white Gaussian noise (AWGN) [3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14]. Though these methods are effective at Gaussian noise removal, their performance on denoising real-world noisy images has seldomly been tested. How to evaluate the performance of these methods is still an open problem.

Over the last decade, several methods [16, 17, 18, 19, 15, 20, 21] are proposed to deal with real-world noisy images. They all employ a two-stage framework for solving the problem of real-world noisy image denoising. In the first stage, these methods assume a model on the noise distribution and estimate the parameters of the model. In the second stage, they perform denoising with the help of the noise modeling and estimation in the first stage. The distribution of noise in real-world noisy images is commonly assumed to be Gaussian distributed [16, 17, 18, 15, 21], mixed Gaussian and Laplacian [19], mixture of Gaussians [20], etc. Although additive white Gaussian is a commonly used noise model by image denoising methods, it is inflexible in describing the complex noise on real-world noisy images [18, 15]. The real camera noise is far beyond Gaussian distributed, signal dependent, and usually hard to estimate [21]. Recently, the mixture of Gaussians (MoG) model is employed to approximate unknown noise in blind image denoising [20]. However, estimating the parameters the MoG model via nonparametric Bayesian techniques [20] or Expectation-Maximization algorithm [23] is usually time consuming [22].

Motivated by the seminar work of [1] that, in the camera imaging process, noise is from several main sources, we

Therefore, how to perform image denoising **without** noise modeling and estimation is an essential while challenging problem. Hence, it is still desirable to design an robust and effective model for real image denoising, which bases on few assumptions and can deal with unknown noise without any parameter tuning procedure. The in-camera imaging pipeline includes image demosaicing, white balance and color space transform, gamut mapping, tone mapping, and JPEG compression. Finally, the major noise in the real image can be categorized into five different types: fixed pattern noise, dark current noise, short noise, amplifier noise, and quantization noise [24].

In this paper, we avoid the challenge problem of noise estimation. We therefore propose a cross domain synthesis solution for real image denoising. In fact, we propose a novel double semi-coupled dictionary learning algorithm for real image denoising problem. In the training stage, given the

108 training patches (noisy ones and clean ones), the dictionaries
109 and coefficients of both the clean and noisy patches. The
110 mapping between the clean and noisy coefficients matrices
111 are also learned in our model. Given a real noisy image,
112 we first extract overlapping patches from it. Then we obtain
113 the clean coefficient from an optimization framework simi-
114 lar to the training stage. The recovered patches are recon-
115 structed by the obtained coefficients on corresponding clean
116 dictionary atoms. We perform comprehensive experiments
117 on real noisy images from multiple different CMOS or CCS
118 sensors. The results demonstrate that our method achieves
119 comparable or even better denoising performance (PSNR,
120 SSIM, and visual quality) on most real noisy images. This
121 reveals that the proposed method has the substantial effect
122 of cross domain image synthesis framework for real image
123 denoising task.

125 1.1. Our Contributions

126 To summarize, the contributions of this paper are as fol-
127 lows:

- 128 • To the best of our knowledge, we are among the first
129 attempts for real image denoising which regard image
130 denoising as a cross domain transfer problem.
- 131 • We also propose a new coupled dictionary learning
132 framework for image restoration problems.
- 133 • We demonstrate that our method achieves the state-of-
134 the-art performance on real image denoising problem,
135 both on objective and subjective measurements.
- 136 • We construct paired dataset by transforming the un-
137 paired dataset via k-Nearest Neighbor algorithm [?].
- 138 • We introduce the Gating Network to speed up the
139 model selection and overall testing speed.

140 2. Related Work

141 2.1. Couple dictionary learning

142 Coupled dictionary learning (CDL) is frequently used in
143 cross-style image synthesis problems such as image super-
144 resolution. CDL assumes that the source and target styles
145 of image have close relationships. CDL aims at learning
146 a pair of dictionaries as well as the relationships between
147 the two cross-domain image styles. Hence, the information
148 from the source image style can be applied to synthesize
149 the image at the target style. The relationships are often as-
150 sumed to be identical mapping (coupled) [25], linear map-
151 ping (semi-coupled) [26]. Yang et al. [25] assumed that LR
152 image patches have the same sparse representations as their
153 HR versions do, and proposed a joint dictionary learning
154 model for SR using concatenated HR/LR image features.

155 They later imposed relaxed constraints on the observed dic-
156 tionary/coefficient pairs across image domains for improved
157 performance. Wang et al. [26] further proposed a semi-
158 coupled dictionary learning (SCDL) scheme by advancing
159 a linear mapping for cross-domain image sparse representa-
160 tion. Their method has been successfully applied to appli-
161 cations of image SR and cross-style synthesis.

162 2.2. Real Image Denoising

163 To the best of our knowledge, the study of real image
164 denoising can be dated back to the BLS-GSM model [27],
165 in which Portilla et al. proposed to use scale mixture of
166 Gaussian in overcomplete oriented pyramids to estimate the
167 latent clean images. In [16], Portilla proposed to use a cor-
168 related Gaussian model for noise estimation of each wavelet
169 subband. Based on the robust statistics theory [?], the work
170 of Rabie [17] modeled the noisy pixels as outliers, which
171 could be removed via Lorentzian robust estimator. In [18],
172 Liu et al. proposed to use 'noise level function' (NLF) to es-
173 timate the noise and then use Gaussian conditional random
174 field to obtain the latent clean image. Recently, Gong et al.
175 proposed an optimization based method [19], which mod-
176 els the data fitting term by weighted sum of ℓ_1 and ℓ_2 norms
177 and the regularization term by sparsity prior in the wavelet
178 transform domain. Later, Lebrun el al. proposed a multi-
179 scale denoising algorithm called 'Noise Clinic' [15] for
180 real image denoising task. This method generalizes the NL-
181 Bayes [28] to deal with signal, scale, and frequency depen-
182 dent noise. Recently, Zhu et al. proposed a Bayesian model
183 [20] which approximates the noise via Mixture of Gaussian
184 (MoG) model [22]. The clean image is recovered from the
185 noisy image by the proposed Low Rank MoG filter (LR-
186 MoG). However, noise level estimation is already a chal-
187 lenging problem and denoising methods are quite sensitive
188 to this parameter. Moreover, these methods are based on
189 shrinkage models that are too simple to reflect reality, which
190 results in over-smoothing of important structures such as
191 small-scale text and textures.

192 3. Double Semi-Couple Dictionary Learning

193 In this section, we first formulate the real image denois-
194 ing problem from the perspective of learning based model
195 and then provide the optimization for the problem.

196 3.1. Problem Formulation

197 For real image denoising, we first collect clean natural
198 images and real noisy images for training. Assume the
199 \mathbf{X} and \mathbf{Y} are unpaired clean image patches and real noisy
200 patches. Let the $\mathbf{X} = \mathbf{D}_x \mathbf{A}_x$ and $\mathbf{Y} = \mathbf{D}_y \mathbf{A}_y + \mathbf{V}_y$, where
201 \mathbf{V}_y is the real noise of which we don't know the distribu-

216 tion.

$$\begin{aligned} & \min_{\mathbf{D}_x, \mathbf{D}_y, \mathbf{A}_x, \mathbf{A}_y} E_{data}(\mathbf{X}, \mathbf{D}_x, \mathbf{A}_x) + E_{data}(\mathbf{Y}, \mathbf{D}_y, \mathbf{A}_y, \mathbf{V}_y) \\ & + E_{map}(f_1(\mathbf{A}_x), f_2(\mathbf{A}_y)) + E_{reg}(\mathbf{A}_x, \mathbf{A}_y, f_1, f_2, \mathbf{D}_x, \mathbf{D}_y, \mathbf{V}_y) \end{aligned} \quad (1)$$

This framework doesn't need noise modeling and estimation. However, we still model the noise by \mathbf{V}_y for visualization what we have removed during training. The regularization of the noise by $\|\mathbf{V}_y\|_p^p$ can be flexible, that we can penalize it by Frobenius norm, ℓ_1 norm, or any other norms. We employ Frobenius norm here for modeling simplicity. To model the relationship between the representational coefficients, we propose to use two invertible mapping function f_1 and f_2 . To measure the error, we employ a penalty function F .

$$\begin{aligned} & \min_{\mathbf{D}_x, \mathbf{D}_y, \mathbf{A}_x, \mathbf{A}_y, \mathbf{U}_x, \mathbf{U}_y, \mathbf{V}_y} \|\mathbf{X} - \mathbf{D}_x \mathbf{A}_x\|_F^2 \\ & + \|\mathbf{Y} - \mathbf{D}_y \mathbf{A}_y - \mathbf{V}_y\|_F^2 + \alpha F(f_1(\mathbf{A}_x), f_2(\mathbf{A}_y)) \\ & + \beta_{x1}\|\mathbf{A}_x\|_1 + \beta_{x2}\|\mathbf{A}_x\|_F^2 + \beta_{y1}\|\mathbf{A}_y\|_1 + \beta_{y2}\|\mathbf{A}_y\|_F^2 \\ & \quad (+\gamma_y\|\mathbf{V}_y\|_p^p) \\ & \text{s.t. } \|\mathbf{d}_{x,i}\|_2 = 1, \|\mathbf{d}_{y,i}\|_2 = 1, \forall i. \end{aligned} \quad (2)$$

Here, we want to discuss more on the mapping functions f_1, f_2 and the measure function F . The mapping function can be linear or nonlinear transformations. The linear function can be defined as a mapping matrix $f_1(\mathbf{A}_x) = \mathbf{U}_x \mathbf{A}_x$ and $f_2(\mathbf{A}_y) = \mathbf{U}_y \mathbf{A}_y$. The corresponding penalty terms on the mapping matrices are $\|\mathbf{U}_x\|_F^2$ and $\|\mathbf{U}_y\|_F^2$. The nonlinear function can be defined as sigmoid function $f_1(\mathbf{A}_x) = 1/(1 + \exp\{-\mathbf{A}_x\})$. We can also employ "first-linear-then-nonlinear" or "first-nonlinear-then-linear" strategies. Here, we don't have explicit penalty terms for the nonlinear mapping functions. The derivatives of the nonlinear case also need further discussions since it is not easy to obtain closed-form solutions with sigmoid functions. In this paper, we utilize linear transformation matrices as the mapping functions f_1 and f_2 . The measure penalty function is simply defined by Frobenius norm. Hence, the term is defined as $\|\mathbf{U}_x \mathbf{A}_x - \mathbf{U}_y \mathbf{A}_y\|_F^2$. However, this would generate a trivial solution of $\mathbf{U}_x = \mathbf{U}_y = \mathbf{0}$. In order to avoid this case, we propose to use the inverse of the mapping matrices, i.e., \mathbf{U}_x^{-1} and \mathbf{U}_y^{-1} .

In summary, we propose a Doubly Inversible and Semi-Coupled Dictionary Learing (DISCDL) model to learn the dictionaries and mapping functions between real noisy im-

ages and latent clean natural images.

$$\begin{aligned} & \min_{\mathbf{D}_x, \mathbf{D}_y, \mathbf{A}_x, \mathbf{A}_y, \mathbf{U}_x, \mathbf{U}_y, \mathbf{V}_y} \|\mathbf{X} - \mathbf{D}_x \mathbf{A}_x\|_F^2 \\ & + \|\mathbf{Y} - \mathbf{D}_y \mathbf{A}_y - \mathbf{V}_y\|_F^2 + \alpha \|\mathbf{U}_x^{-1} \mathbf{A}_x - \mathbf{U}_y^{-1} \mathbf{A}_y\|_F^2 \\ & + \beta_{x1}\|\mathbf{A}_x\|_1 + \beta_{x2}\|\mathbf{A}_x\|_F^2 + \beta_{y1}\|\mathbf{A}_y\|_1 + \beta_{y2}\|\mathbf{A}_y\|_F^2 \\ & \quad (+\gamma_y\|\mathbf{V}_y\|_p^p) \\ & + \lambda_x\|\mathbf{U}_x^{-1}\|_F^2 + \lambda_y\|\mathbf{U}_y^{-1}\|_F^2 \\ & \text{s.t. } \|\mathbf{d}_{x,i}\|_2 = 1, \|\mathbf{d}_{y,i}\|_2 = 1, \forall i. \end{aligned} \quad (3)$$

This model has three major differences when compared with SCDL model.

- We use a matrix \mathbf{V}_y to model the noise, and we don't set any prior distribution on it. This term can help us visualize the noise we learned from the data, i.e., the real noisy images. This make our model fully data-driven. Since our assumption (we have no assumption at all) on noise is more flexible than others', the noise we obtain in our model can be more accurate than other statistical models such as Gaussian or Mixture of Gaussians. Besides, it is time-consuming to fit the noise model from the online data.
- We use two invertible matrices as the mapping transformations between the coefficients of the real noisy patches and the latent clean patches. This makes our model more flexible than SCDL in which the mapping matrix not explicitly invertible. Besides, the SCDL can only transform LR images into HG images while our model can transform two different image styles in both direction.
- The constraints on dictionary atoms in our model is strictly $\|\mathbf{d}_{x,i}\|_2 = 1, \|\mathbf{d}_{y,i}\|_2 = 1$ while the CDL model and SCDL model are $\|\mathbf{d}_{x,i}\|_2 \leq 1, \|\mathbf{d}_{y,i}\|_2 \leq 1$. This makes our model more robust on the dictionary learning since both the dictionary atoms and sparse coefficients are interacted with each other. The ≤ 1 constraints would like to make the coefficients larger and dictionary atoms smaller or even vanish. However, in the training stage, we care more about the dictionary atoms and would rather ignore the sparse coefficients.

3.2. Model Optimization

While the objective function in (3) is not convex, it is convex with each variable when other variables are fixed. We employ alternating direction method of multipliers (ADMM) algorithm here. Specifically, we divide the objective function into four sub-problems: 1) updating the sparse coefficients $\mathbf{A}_x, \mathbf{A}_y$; 2) updating the normalized dictionaries $\mathbf{D}_x, \mathbf{D}_y$; 3) updating the noise matrix \mathbf{V}_y ; 4) updating the mapping matirces $\mathbf{U}_x, \mathbf{U}_y$. We discuss the four steps as follows.

324 **3.2.1 Updating \mathbf{A}_x and \mathbf{A}_y**

325
$$\min_{\mathbf{A}_x} \|\mathbf{X} - \mathbf{D}_x \mathbf{A}_x\|_F^2 + \alpha \|\mathbf{U}_x^{-1} \mathbf{A}_x - \mathbf{U}_y^{-1} \mathbf{A}_y\|_F^2 + \beta_{x1} \|\mathbf{A}_x\|_1 + \beta_{x2} \|\mathbf{A}_x\|_F^2, \quad (4)$$

326
$$\min_{\mathbf{A}_y} \|\mathbf{Y} - \mathbf{D}_y \mathbf{A}_y - \mathbf{V}_y\|_F^2 + \alpha \|\mathbf{U}_x^{-1} \mathbf{A}_x - \mathbf{U}_y^{-1} \mathbf{A}_y\|_F^2 + \beta_{y1} \|\mathbf{A}_y\|_1 + \beta_{y2} \|\mathbf{A}_y\|_F^2. \quad (5)$$

334 Take \mathbf{A}_x as an example, the first and second terms above
335 can be combined to form a new optimization problems as
336 follows:

338
$$\min_{\mathbf{A}_x} \|\tilde{\mathbf{X}} - \tilde{\mathbf{D}}_x \mathbf{A}_x\|_F^2 + \beta_{x1} \|\mathbf{A}_x\|_1 + \beta_{x2} \|\mathbf{A}_x\|_F^2, \quad (6)$$

340 where $\tilde{\mathbf{X}} = \begin{pmatrix} \mathbf{X} \\ \sqrt{\alpha} \mathbf{U}_y^{-1} \mathbf{A}_y \end{pmatrix}$ and $\tilde{\mathbf{D}} = \begin{pmatrix} \mathbf{D}_x \\ \sqrt{\alpha} \mathbf{U}_x^{-1} \end{pmatrix}$.
341 For \mathbf{A}_y , it is similar with \mathbf{A}_x .

344
$$\min_{\mathbf{A}_y} \|\tilde{\mathbf{Y}} - \tilde{\mathbf{D}}_y \mathbf{A}_y\|_F^2 + \beta_{y1} \|\mathbf{A}_y\|_1 + \beta_{y2} \|\mathbf{A}_y\|_F^2, \quad (7)$$

347 where $\tilde{\mathbf{Y}} = \begin{pmatrix} \mathbf{Y} - \mathbf{V}_y \\ \sqrt{\alpha} \mathbf{U}_x^{-1} \mathbf{A}_x \end{pmatrix}$ and $\tilde{\mathbf{D}} = \begin{pmatrix} \mathbf{D}_y \\ \sqrt{\alpha} \mathbf{U}_y^{-1} \end{pmatrix}$.
348 These simplified versions have the exactly same formulation
349 as standard sparse coding and can be simply solved by
350 tools such as SPAMS.

352 The \mathbf{U}_x^{-1} and \mathbf{U}_y^{-1} are invertible. This will be discussed
353 in subsection "Updating U".

3.2.2 Updating \mathbf{D}_x and \mathbf{D}_y

357
$$\min_{\mathbf{D}_x} \|\mathbf{X} - \mathbf{D}_x \mathbf{A}_x\|_F^2 \quad \text{s.t.} \quad \|\mathbf{d}_{x,i}\|_2 = 1, \forall i. \quad (8)$$

360
$$\min_{\mathbf{D}_y} \|\mathbf{Y} - \mathbf{D}_y \mathbf{A}_y - \mathbf{V}_y\|_F^2 \quad \text{s.t.} \quad \|\mathbf{d}_{y,i}\|_2 = 1, \forall i. \quad (9)$$

363 These two are quadraically constrained quadratic program
364 (QCQP) problem and can be solved by Lagrange dual tech-
365 niques.

3.2.3 Updating \mathbf{V}_y

369 The noise matrix is initialized as a zero matirx and updated
370 by solving the following probelm:

372
$$\min_{\mathbf{V}_y} \|\mathbf{Y} - \mathbf{D}_y \mathbf{A}_y - \mathbf{V}_y\|_F^2 + \gamma_y \|\mathbf{V}_y\|_F^2 \quad (10)$$

374 This is a ridge regression problem. We can obtain the ana-
375 lytical solution of \mathbf{V}_y by

377
$$\mathbf{V}_y = (\mathbf{Y} - \mathbf{D}_y \mathbf{A}_y) / (1 + \gamma_y). \quad (11)$$

3.2.4 Alternate Updating \mathbf{V}_y

The noise matrix is initialized as a zero matirx and updated
by solving the following probelm:

$$\min_{\mathbf{V}_y} \|\mathbf{Y} - \mathbf{D}_y \mathbf{A}_y - \mathbf{V}_y\|_F^2 \quad (12)$$

This is a standard least square problem. We can obtain the analytical solution of \mathbf{V}_y by

$$\mathbf{V}_y = \mathbf{Y} - \mathbf{D}_y \mathbf{A}_y. \quad (13)$$

3.2.5 Updating \mathbf{U}_x and \mathbf{U}_y

$$\begin{aligned} & \min_{\mathbf{U}_x^{-1}} \alpha \|\mathbf{U}_y^{-1} \mathbf{A}_y - \mathbf{U}_x^{-1} \mathbf{A}_x\|_F^2 + \lambda_x \|\mathbf{U}_x^{-1}\|_F^2 \\ & \min_{\mathbf{U}_y^{-1}} \alpha \|\mathbf{U}_x^{-1} \mathbf{A}_x - \mathbf{U}_y^{-1} \mathbf{A}_y\|_F^2 + \lambda_y \|\mathbf{U}_y^{-1}\|_F^2 \end{aligned} \quad (14)$$

The above problems are also ridge regression problems and have analytical solutions of \mathbf{U}_x and \mathbf{U}_y as follows:

$$\begin{aligned} \mathbf{U}_x^{-1} &= \mathbf{U}_y^{-1} \mathbf{A}_y \mathbf{A}_x^T (\mathbf{A}_x \mathbf{A}_x^T + (\gamma_x/\alpha) \mathbf{I})^{-1} \\ \mathbf{U}_y^{-1} &= \mathbf{U}_x^{-1} \mathbf{A}_x \mathbf{A}_y^T (\mathbf{A}_y \mathbf{A}_y^T + (\gamma_y/\alpha) \mathbf{I})^{-1} \end{aligned} \quad (15)$$

Here, we verify that \mathbf{U}_x^{-1} and \mathbf{U}_y^{-1} are invertible. The \mathbf{U}_x^{-1} and \mathbf{U}_y^{-1} are both initialized as an identity matrix, of suitable dimension, which is inversible. That is, we have $\mathbf{U}_y^{(0)} = \mathbf{I}$ when we compute \mathbf{U}_x^{-1} . If $\mathbf{A}_y \mathbf{A}_x^T$ is inversible, then \mathbf{U}_x^{-1} is inversible. In fact, we have $\mathbf{A}_y, \mathbf{A}_x \in \mathbb{R}^{d \times N}$. d is the dimension of the sample. For a patch of size 8×8 , $d = 64$. The N is the number of samples in the training data. Remember that we have much more samples when compared to the dimension of patches, that is $N \gg d$. It is less likely that $\mathbf{A}_y \mathbf{A}_x^T \in \mathbb{R}^{d \times d}$ has a rank structure lower than d . In other words, $\mathbf{A}_y \mathbf{A}_x^T \in \mathbb{R}^{d \times d}$ is less likely to be singular if we have enough training data. The experiments also confirm our conjecture. Besides, we can also add small disburcation to guarantee that $\mathbf{A}_y \mathbf{A}_x^T \in \mathbb{R}^{d \times d}$ is inversible.

Once \mathbf{U}_x^{-1} is inversible, we can also verify that \mathbf{U}_y^{-1} is inversible in a similar way.

3.3 Real Image Denoising

Two methods:

The first one is that

$$\begin{aligned} & \min_{\mathbf{a}_{x,i}, \mathbf{a}_{y,i}} \|\mathbf{x}_i - \mathbf{D}_x \mathbf{a}_{x,i}\|_2^2 + \|\mathbf{y}_i - \mathbf{D}_y \mathbf{a}_{y,i} - \mathbf{v}_{y,i}\|_2^2 \\ & + \alpha \|\mathbf{U}_x^{-1} \mathbf{a}_{x,i} - \mathbf{U}_y^{-1} \mathbf{a}_{y,i}\|_2^2 \\ & + \beta_x \|\mathbf{a}_{x,i}\|_1 + \beta_{x2} \|\mathbf{a}_{x,i}\|_2^2 + \beta_y \|\mathbf{a}_{y,i}\|_1 + \beta_{y2} \|\mathbf{a}_{y,i}\|_2^2 \\ & (+\gamma_y \|\mathbf{v}_{y,i}\|_1) \end{aligned} \quad (16)$$

and finally we get $\hat{\mathbf{x}}_i = \mathbf{D}_x \hat{\mathbf{a}}_{x,i}$.

432 The second one is to solve
 433

$$\begin{aligned} \min_{\mathbf{a}_{y,i}, \mathbf{v}_{y,i}} & \|\mathbf{y}_i - \mathbf{D}_y \mathbf{a}_{y,i} - \mathbf{v}_{y,i}\|_2^2 + \alpha \|\mathbf{U}_x^{-1} \mathbf{a}_{x,i} - \mathbf{U}_y^{-1} \mathbf{a}_{y,i}\|_2^2 \\ & + \beta_{y1} \|\mathbf{a}_{y,i}\|_1 + \beta_{y2} \|\mathbf{a}_{y,i}\|_2^2 \\ & (+\gamma_y \|\mathbf{v}_{y,i}\|_1) \end{aligned} \quad (17)$$

439 Once we get $\hat{\mathbf{a}}_{y,i}$ from \mathbf{y}_i , $\hat{\mathbf{a}}_{x,i} \approx \mathbf{U}_x \mathbf{U}_y^{-1} \hat{\mathbf{a}}_{y,i}$ and $\hat{\mathbf{x}}_i \approx \mathbf{D}_x \hat{\mathbf{a}}_{x,i}$.
 440

441 Experiments demonstrate that the first method can get
 442 better performance than the second one while the second
 443 one can get faster speed than the first one.
 444

445 We can also initialized the solution from the second one.
 446

4. The Overall Algorithm

4.1. Pair Sample Construction from Unpaired Samples

451 In cross style transfer methods such as CDL and SCDL,
 452 the authors assume that the two different styles have paired
 453 data, i.e., for each data sample in one style, we can find
 454 paired data sample in the other style. However, in real
 455 world, the data from two different sources may be un-
 456 paired. For example, the real noisy images should not
 457 have groundtruth clean images of the same scene. The real
 458 low-resolution images should not have corresponding high-
 459 resolution images in the real world. The real blurry images
 460 should not have corresponding clear and high quality im-
 461 ages in real world.

462 To deal with unpaired data, we could collect real noisy
 463 images and clean natural images from two different sources.
 464 The real noisy images are from the example images (18 im-
 465 ages) of the Neat Image website while the clean natural im-
 466 ages are from the training set (200 images) of the Berkeley
 467 Segmentation Dataset (BSDS500). To make use of the un-
 468 paired data samples, we employ searching strategy to con-
 469 struct the training dataset. That is, for each noisy image
 470 patch, we utilize the k-Nearest Neighbor (k-NN) algorithm
 471 to find the most similar patch in the clean images as the
 472 paired groundtruth patch. The similarity is measured by the
 473 Euclidean distance (also called squared error or ℓ_2 norm).

4.2. Structural Clustering and Model Selection

474 In fact, different image structures should have differ-
 475 ent influences on dictioanry as well as the mapping func-
 476 tion. Patches with flat region should have low rank struc-
 477 ture within dictionary elements and identity mapping be-
 478 tween noisy and latent clean patches. Patches with com-
 479 plex details should have more comprehensive dictionary el-
 480 ements within dictionary elements and more complex map-
 481 ping function between noisy and clean patches. A single
 482 mapping function cannot deal with all these complex rela-
 483 tionships. Hence, a structural clustering procedure is needed
 484

485 for complex solution. In this paper, we propose to em-
 486 ploy Gaussian Mixture Model to cluster different image
 487 patches into different groups and learn dictionary and map-
 488 ping function for each group.
 489

4.3. Adaptive Iterations of Different Noise Levels

490 For real image denoising, we can perform well on images
 491 which have similar noise levels with the training dataset.
 492 How can we deal with the real noisy images whose noise
 493 levels are higher than the training dataset? The answer is
 494 to remove the noise by more iterations. The input image of
 495 each iteration is the recovered image of previous iteration.
 496 This makes sense since we can still view the recovered im-
 497 age as a real noisy image.
 498

499 This will also bring a second problem, that how we could
 500 automatically terminate the iteration. This can be solved
 501 by two methods. One way is to compare the images be-
 502 tween two iterations and calculate their difference, the it-
 503 eration can be terminated if the difference is smaller than
 504 a threshold. The other way is to estimate the noise level
 505 of the current image and terminate the iterations when the
 506 noise level is lower than a preset threshold. We employ the
 507 second way and set the threshold as 0.0001 in our experi-
 508 ments. In fact, most of our testing images will be denoised
 509 well in one iteration.
 510

4.4. Efficient Model Selection by Gating Network

511 In the Gaussian component selection procedure, if we
 512 employ the full posterior estimation, the speed is not fast.
 513 Our algorithm can be speeded up by introducing the Gating
 514 network model.
 515

5. Experiments

516 We compare with popular software NeatImage which
 517 is one of the best denoising software available. All these
 518 methods need noise estimation which is vary hard to per-
 519 form if there is no uniform regions are available in the test-
 520 ing image. The NeatImage will fail to perform automatical-
 521 ly parameters settings if there is no uniform regions.
 522

5.1. Parameters

523 We don't fine tune the parameters both in the training
 524 and testing datasets.
 525

5.2. Real Image Denoising

526 We compare the proposed method with the famous
 527 BM3D [7] and WNNM [11], Cascade of Shrinkage Fields
 528 (CSF) [12], trainable reaction diffusion (TRD) [13], plain
 529 neural network based method MLP [14], the blind image
 530 denoising method Noise Clinic [15], and the commercial
 531 software Neat Image. The RGB images are firstly trans-
 532 formed into YCbCr channels and restored by these methods.
 533

540 Then the denoised RGB image is obtained by transforming
 541 the restored YCbCr image back.
 542

543 We evaluate the competing denoising methods from various
 544 research directions on two datasets. Both the two
 545 datasets comes from the [21]. The first contains 3 cropped
 546 images of size 512×512 . The other dataset contains 42
 547 images cropped to size of 500×500 from the 17 images
 548 provided in [21]. The 60 images contain most of the scenes
 549 in the 17 images [21].

550 6. Conclusion and Future Work

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

560 References

- [1] 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
- [2] P. Arbelaez, M. Maire, C. Fowlkes, and J. Malik. Contour detection and hierarchical image segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.*, 33(5):898–916, 2011.
- [3] L. I. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D: Nonlinear Phenomena*, 60(1):259–268, 1992. 1
- [4] A. Buades, B. Coll, and J. M. Morel. A non-local algorithm for image denoising. *CVPR*, pages 60–65, 2005. 1
- [5] S. Roth and M. J. Black. Fields of Experts. *International Journal of Computer Vision*, 82(2):205–229, 2009. 1
- [6] 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
- [7] 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, 5
- [8] J. Mairal, F. Bach, J. Ponce, G. Sapiro, and A. Zisserman. Non-local sparse models for image restoration. *ICCV*, pages 2272–2279, 2009. 1
- [9] D. Zoran and Y. Weiss. From learning models of natural image patches to whole image restoration. *ICCV*, pages 479–486, 2011. 1
- [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
- [11] S. Gu, L. Zhang, W. Zuo, and X. Feng. Weighted nuclear norm minimization with application to image denoising. *CVPR*, pages 2862–2869, 2014. 1, 5
- [12] 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, 5
- [13] 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, 5
- [14] 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, 5
- [15] M. Lebrun, M. Colom, and J.-M. Morel. Multiscale image blind denoising. *Image Processing, IEEE Transactions on*, 24(10):3149–3161, 2015. 1, 2, 5
- [16] 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, 2
- [17] Tamer Rabie. Robust estimation approach for blind denoising. *Image Processing, IEEE Transactions on*, 14(11):1755–1765, 2005. 1, 2
- [18] 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
- [19] Zheng Gong, Zuowei Shen, and Kim-Chuan Toh. Image restoration with mixed or unknown noises. *Multiscale Modeling & Simulation*, 12(2):458–487, 2014. 1, 2
- [20] 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, 2
- [21] 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, 6, 7, 8
- [22] C. M. Bishop. *Pattern recognition and machine learning*. New York: Springer, 2006. 1, 2
- [23] 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. 1
- [24] Yanghai Tsin, Visvanathan Ramesh, and Takeo Kanade. Statistical calibration of ccd imaging process. *Computer Vision, 2001. ICCV 2001. Proceedings. Eighth IEEE International Conference on*, 1:480–487, 2001. 1

648 Table 1. Average PSNR(dB) results of different methods on 3 real noisy images captured by Canon EOS 5D mark3 at ISO3200 in [21]. 702

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	Image	Noisy	BM3D	WNNM	CSF	TRD	MLP	Noise Clinic	Neat Image	Ours
	1	37.00	37.08	37.09	37.46	37.51	32.91	38.76	37.68	38.63
	2	33.88	33.95	33.95	34.90	35.04	31.94	35.69	34.87	35.96
	3	33.83	33.85	33.85	34.15	34.07	30.89	35.54	34.77	35.51
	Average	34.90	34.96	34.96	35.50	35.54	31.91	36.67	35.77	36.70

656 Table 2. Average SSIM results of different methods on 3 real noisy images captured by Canon EOS 5D mark3 at ISO3200 in [21]. 710

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	Image	Noisy	BM3D	WNNM	CSF	TRD	MLP	Noise Clinic	Neat Image	Ours
	1	0.9345	0.9368	0.9372	0.9599	0.9607	0.9043	0.9689	0.9600	0.9712
	2	0.8919	0.8848	0.8951	0.9159	0.9187	0.8498	0.9427	0.9308	0.9434
	3	0.9128	0.9136	0.9136	0.9254	0.9279	0.8635	0.9476	0.9463	0.9529
	Average	0.9131	0.9117	0.9153	0.9337	0.9358	0.8725	0.9531	0.9457	0.9558

- [25] Jianchao Yang, John Wright, Thomas S Huang, and Yi Ma. Image super-resolution via sparse representation. *IEEE transactions on image processing*, 19(11):2861–2873, 2010. 2
- [26] Shenlong Wang, Lei Zhang, Yan Liang, and Quan Pan. Semi-coupled dictionary learning with applications to image super-resolution and photo-sketch synthesis. *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, pages 2216–2223, 2012. 2
- [27] J. Portilla, V. Strela, M.J. Wainwright, and E.P. Simoncelli. Image denoising using scale mixtures of Gaussians in the wavelet domain. *Image Processing, IEEE Transactions on*, 12(11):1338–1351, 2003. 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. 2

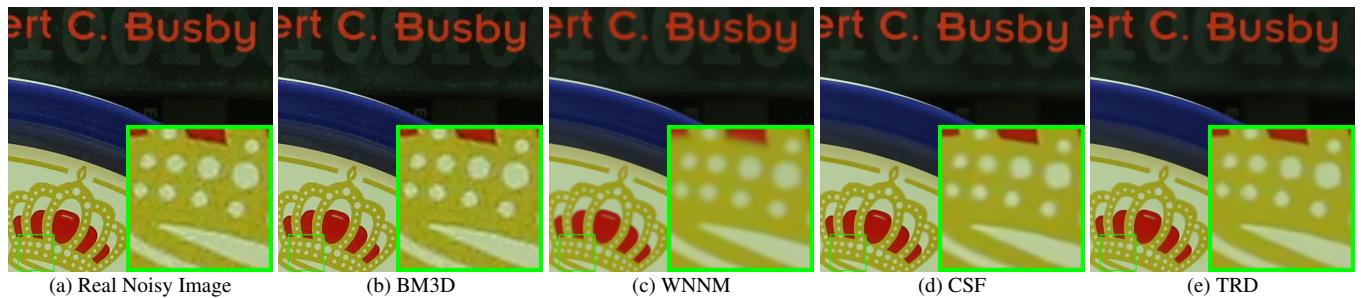
756
757
758
759
760
761
762
763
764
765
766810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863

Figure 1. Denoised images of the old image "5dmark3iso32001" by different methods. The images are better to be zoomed in on screen.

777

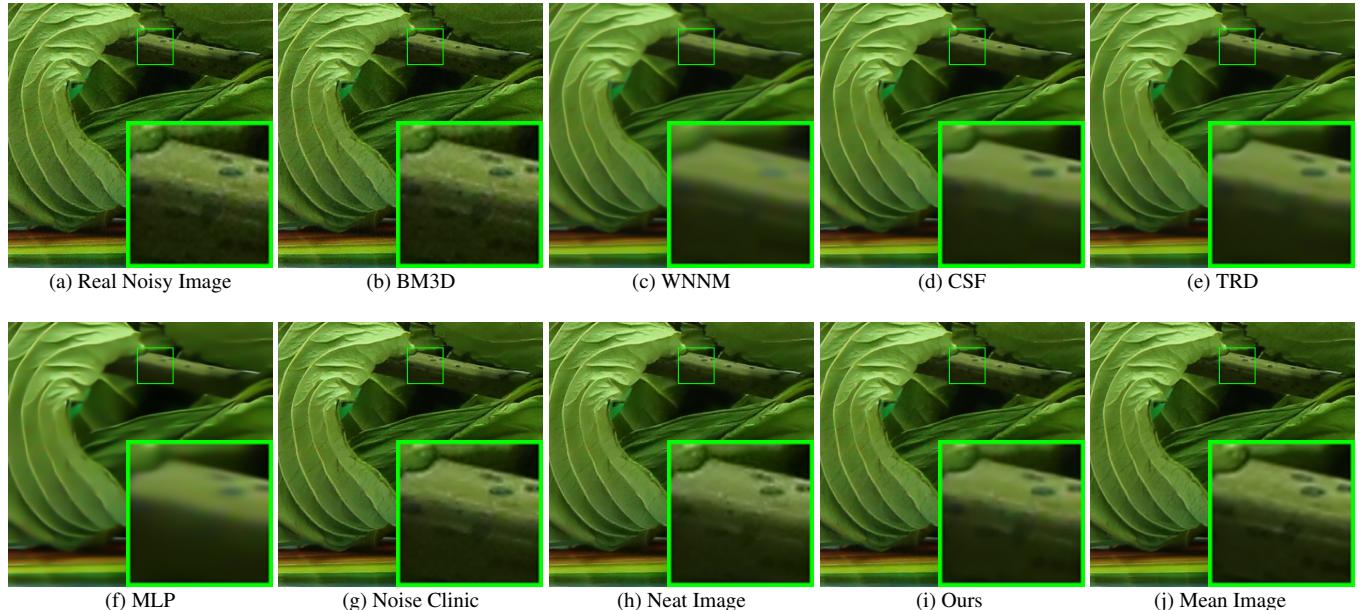


Figure 2. Denoised images of the old image "5dmark3iso32002" by different methods. The images are better to be zoomed in on screen.

801

802

803

Table 3. Average PSNR(dB) and SSIM results of different methods on 42 cropped images from 17 real noisy images in [21].

Measure	Noisy	BM3D	WNNM	CSF	TRD	MLP	Noise Clinic	Neat Image	Ours
PSNR	34.36	34.36	34.40	36.11	36.05	34.41	37.68	36.58	36.15
SSIM	0.8552	0.8553	0.8577	0.9215	0.9211	0.9012	0.9470	0.9145	0.9236

864
865
866
867
868
869
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

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

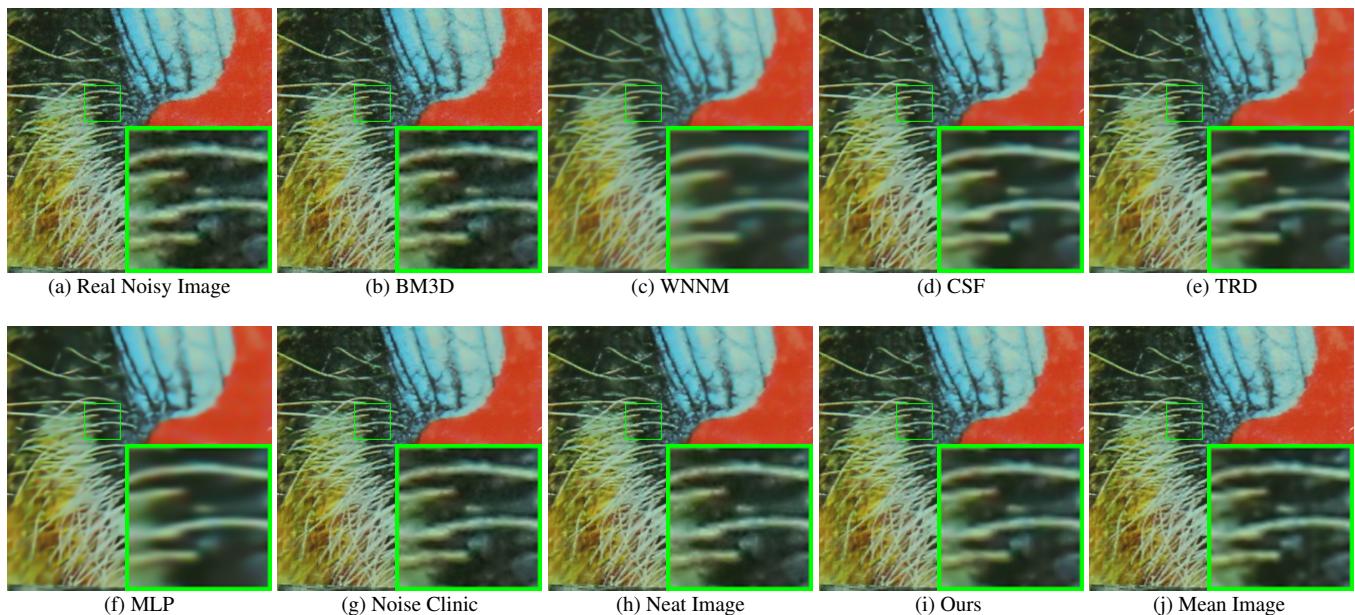


Figure 3. Denoised images of the old image "5dmark3iso3200_3" by different methods. The images are better to be zoomed in on screen.