

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) [2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12]. Though these methods are effective at Gaussian noise removal, their performance on denoising real-world noisy images has seldomly been tested.

Over the last decade, several methods [13, 14, 15, 16,

17, 18, 19] are proposed to deal with real-world noisy images. These methods coincidentally employ a two-stage framework. 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 [13, 14, 15, 17, 19], mixed Gaussian and Laplacian [16], mixture of Gaussians [18], 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 [15, 17]. The real camera noise is far beyond Gaussian distributed, signal dependent, and usually hard to estimate [19]. Recently, the mixture of Gaussians (MoG) model is employed to approximate unknown noise in blind image denoising [18]. However, estimating the parameters the MoG model via nonparametric Bayesian techniques [18] or Expectation-Maximization algorithm [20] is usually time consuming [21].

To avoid the latent problems, it is naturally to ask, whether it is possible to perform denoising on real-world noisy images without noise modeling and estimation? To answer this question, we first look closer into several learning based methods working effectively on Gaussian noise removal. Given clean images and the noisy counterparts degraded by identical Gaussian noise, these methods trained a plain neural network [8], cascades of shrinkage filters [10], or a reaction diffusion process [12] for image denoising. These methods are effective on denoising the images being degraded in the same way, i.e., Gaussian noise with zero mean and fixed variance, as in the training stage. However, the noise in real-world noisy images are much more complex than Gaussian.

Almost all existing discriminative learning methods for image enhancement [refs] require a set of paired images, denoted by P: (x_i, y_i) , $i=1,2,N$, to train the desired model, where x_i is a high quality ground-truth image and y_i is its degraded counterpart. For example, in [refs], a set of clean images and their simulated noisy counterparts (mostly additive white Gaussian white noise, AWGN) are used

108 to learn the discriminative denoising model; in [refs], a
109 set of high-resolution images and their down-sampled low-
110 resolution counterparts (by using certain down-sampling
111 operators such as bi-cubic interpolator) are used to learn
112 the super-resolution models. Unfortunately, the learned dis-
113 criminative models by these methods will have big limi-
114 tations in real applications due to the usage of simulated
115 paired data for training. For example, the noise in on
116 real-world noisy images will be much more complex than
117 AWGN, and thus the model learned with AWGN will
118 become much less effective for real-world noisy images. In
119 Fig. 1, we applied the models learned by [refs] to a real-
120 world noisy image captured by a digital camera. One can
121 see that there are still many noise caused artifacts remain-
122 ing in on the denoised images. The above mentioned limi-
123 tations of existing discriminative learning methods motivate
124 us to develop new discriminative learning methods without
125 paired training data, aiming to provide more effective and
126 practical solutions for image enhancement. Instead of sim-
127 ulating the paired images for learning, we propose to use a set
128 of real-world degraded images, denoted by $Y: y_i, i=1,2,..,M$,
129 and an independent set of high quality images, denoted by
130 $X: x_j, j=1,2,..,N$, to train discriminative models for practical
131 image enhancement. Note that in our problem, any image in
132 y_i does not have a counterpart image in x_j . It is much more
133 challenging than the learning problem with paired data, and
134 all the current discriminative learning methods [refs] cannot
135 be applied.

136 Natural images are of high dimensionality and vary sig-
137 nificantly in content, making the learning of statistical mod-
138 els for the whole image very difficult. Considering that im-
139 age patches are building blocks of an image, one commonly
140 used strategy is to learn statistical models for image patches
141 [refs]. This largely reduces the searching space of desired
142 models and makes the learning problem easier to solve. In
143 particular, for our learning problem, though images in X
144 and Y are different, it is possible to find a group of patches
145 in X which are similar to a given group of patches in Y .

146 Hence, it is still desirable to design an robust and effec-
147 tive model for real image denoising, which bases on few as-
148 sumptions and can deal with unknown noise without any pa-
149 rameter tuning procedure. The in-camera imaging pipeline
150 includes image demosaicing, white balance and color space
151 transform, gamut mapping, tone mapping, and JPEG com-
152 pression. Finally, the major noise in the real image can be
153 categorized into five different types: fixed pattern noise,
154 dark current noise, short noise, amplifier noise, and quan-
155 tization noise [22].

156 In this paper, we avoid the challenge problem of noise es-
157 timation. We therefore propose a cross domain synthesis so-
158 lution for real image denoising. In fact, we propose a noval
159 double semi-couplde dictionary learning algorithm for real
160 image denoising problem. In the training stage, given the

162 training patches (niosy ones and clean ones), the dictionar-
163 ries and coefficients of both the clean and noisy patches.The
164 mapping between the clean and noisy coefficients matrices
165 are also learned in our model. Given a real noisy image,
166 we first extract overlapping patches from it. Then we obtain
167 the clean coefficient from an optimization framework sim-
168 ilar to the training stage. The recovered patches are recon-
169 structed by the obtained coefficients on corresponding clean
170 dictionary atoms. We perform comprehensive experiments
171 on real noisy images from multiple different CMOS or CCS
172 sensors. The results demonstrate that our method achieves
173 comparable or even better denoising performance (PSNR,
174 SSIM, and visual quality) on most real noisy images. This
175 reveals that the proposed method has the substantial effect
176 of cross domain image synthesis framework for real image
177 denoising task.

1.1. Our Contributions

To summerize, the contributions of this paper are as fol-
lows:

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

2. Related Work

2.1. Couple dictionary learning

Coupled dictionary learning (CDL) is frequently used in cross-style image synthesis problems such as image super-resolution. CDL assumes that the source and target styles of image have close relationships. CDL aims at learning a pair of dictionaries as well as the relationships between the two cross-domain image styles. Hence, the information from the source image style can be applied to synthesize the image at the target style. The relationships are often assumed to be identical mapping (coupled) [23], linear mapping (semi-coupled) [24]. Yang et al. [23] assumed that LR image patches have the same sparse representations as their HR versions do, and proposed a joint dictionary learning model for SR using concatenated HR/LR image features.

216 They later imposed relaxed constraints on the observed dictionary/coefficients pairs across image domains for improved
 217 performance. Wang et al. [24] further proposed a semi-
 218 coupled dictionary learning (SCDL) scheme by advancing
 219 a linear mapping for cross-domain image sparse representa-
 220 tion. Their method has been successfully applied to applica-
 221 tions of image SR and cross-style synthesis.
 222

224 225 2.2. Real Image Denoising

226 To the best of our knowledge, the study of real image
 227 denoising can be dated back to the BLS-GSM model [25],
 228 in which Portilla et al. proposed to use scale mixture of
 229 Gaussian in overcomplete oriented pyramids to estimate the
 230 latent clean images. In [13], Portilla proposed to use a cor-
 231 related Gaussian model for noise estimation of each wavelet
 232 subband. Based on the robust statistics theory [?], the work
 233 of Rabie [14] modeled the noisy pixels as outliers, which
 234 could be removed via Lorentzian robust estimator. In [15],
 235 Liu et al. proposed to use 'noise level function' (NLF) to es-
 236 timate the noise and then use Gaussian conditional random
 237 field to obtain the latent clean image. Recently, Gong et al.
 238 proposed an optimization based method [16], which mod-
 239 els the data fitting term by weighted sum of ℓ_1 and ℓ_2 norms
 240 and the regularization term by sparsity prior in the wavelet
 241 transform domain. Later, Lebrun et al. proposed a multi-
 242 scale denoising algorithm called 'Noise Clinic' [17] for
 243 real image denoising task. This method generalizes the NL-
 244 Bayes [26] to deal with signal, scale, and frequency depen-
 245 dent noise. Recently, Zhu et al. proposed a Bayesian model
 246 [18] which approximates the noise via Mixture of Gaussian
 247 (MoG) model [21]. The clean image is recovered from the
 248 noisy image by the proposed Low Rank MoG filter (LR-
 249 MoG). However, noise level estimation is already a chal-
 250 lenging problem and denoising methods are quite sensitive
 251 to this parameter. Moreover, these methods are based on
 252 shrinkage models that are too simple to reflect reality, which
 253 results in over-smoothing of important structures such as
 254 small-scale text and textures.
 255

256 257 3. Double Semi-Couple Dictionary Learning

258 In this section, we first formulate the real image denois-
 259 ing problem from the perspective of learning based model
 260 and then provide the optimization for the problem.
 261

262 263 3.1. Problem Formulation

264 For real image denoising, we first collect clean natural
 265 images and real noisy images for training. Assume the
 266 \mathbf{X} and \mathbf{Y} are unpaired clean image patches and real noisy
 267 patches. Let the $\mathbf{X} = \mathbf{D}_x \mathbf{A}_x$ and $\mathbf{Y} = \mathbf{D}_y \mathbf{A}_y + \mathbf{V}_y$, where
 268 \mathbf{V}_y is the real noise of which we don't know the distribu-
 269

270 tion.
 271

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

This framework doesn't need noise modeling and esti-
 280 mation. However, we still model the noise by \mathbf{V}_y for visu-
 281 alization what we have removed during training. The regu-
 282 larization of the noise by $\|\mathbf{V}_y\|_p^p$ can be flexible, that we can
 283 penalize it by Frobenius norm, ℓ_1 norm, or any other norms.
 284 We employ Frobenius norm here for modeling simplicity.
 285 To model the relationship between the representational co-
 286 efficients, we propose to use two invertible mapping func-
 287 tion f_1 and f_2 . To measure the error, we employ a penalty
 288 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 func-
 tion 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 nonlin-
 ear 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 map-
 ping 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 func-
 tions 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 triv-
 ial 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-

324 ages and latent clean natural images.
 325

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

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

- 338 • We use a matrix \mathbf{V}_y to model the noise, and we don't
 339 set any prior distribution on it. This term can help us
 340 visualize the noise we learned from the data, i.e., the
 341 real noisy images. This make our model fully data-
 342 driven. Since our assumption (we have no assump-
 343 tion at all) on noise is more flexible than others', the
 344 noise we obtain in our model can be more accurate
 345 than other statistical models such as Gaussian or Mix-
 346 ture of Gaussians. Besides, it is time-consuming to fit
 347 the noise model from the online data.
- 348 • We use two invertible matrices as the mapping trans-
 349 formations between the coefficients of the real noisy
 350 patches and the latent clean patches. This makes our
 351 model more flexible than SCDL in which the mapping
 352 matrix not explicitly invertible. Besides, the SCDL can
 353 only transform LR images into HG images while our
 354 model can transform two different image styles in both
 355 direction.
- 356 • The constraints on dictionary atoms in our model is
 357 strictly $\|\mathbf{d}_{x,i}\|_2 = 1, \|\mathbf{d}_{y,i}\|_2 = 1$ while the CDL
 358 model and SCDL model are $\|\mathbf{d}_{x,i}\|_2 \leq 1, \|\mathbf{d}_{y,i}\|_2 \leq$
 359 1. This makes our model more robust on the dictionary
 360 learning since both the dictionary atoms and sparse
 361 coefficients are interacted with each other. The ≤ 1
 362 constraints would like to make the coefficients larger and
 363 dictionary atoms smaller or even vanish. However, in
 364 the training stage, we care more about the dictionary
 365 atoms and would rather ignore the sparse coefficients.

367 3.2. Model Optimization

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

378 3.2.1 Updating \mathbf{A}_x and \mathbf{A}_y

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

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

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

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

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

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

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

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

411 3.2.2 Updating \mathbf{D}_x and \mathbf{D}_y

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

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

413 These two are quadraically constrained quadratic program
 414 (QCQP) problem and can be solved by Lagrange dual tech-
 415 niques.

421 3.2.3 Updating \mathbf{V}_y

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

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

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

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

432 **3.2.4 Alternate Updating \mathbf{V}_y** 486

433 The noise matrix is initialized as a zero matirx and updated
 434 by solving the following probelm:
 435

$$436 \quad \min_{\mathbf{V}_y} \|\mathbf{Y} - \mathbf{D}_y \mathbf{A}_y - \mathbf{V}_y\|_F^2 \quad (12)$$

439 This is a standard least square problem. We can obtain the
 440 analytical solution of \mathbf{V}_y by
 441

$$442 \quad \mathbf{V}_y = \mathbf{Y} - \mathbf{D}_y \mathbf{A}_y. \quad (13)$$

444 **3.2.5 Updating \mathbf{U}_x and \mathbf{U}_y** 493

$$446 \quad \begin{aligned} & \min_{\mathbf{U}_x^{-1}} \alpha \|\mathbf{U}_y^{-1} \mathbf{A}_y - \mathbf{U}_x^{-1} \mathbf{A}_x\|_F^2 + \lambda_x \|\mathbf{U}_x^{-1}\|_F^2 \\ & \min_{\mathbf{U}_y^{-1}} \alpha \|\mathbf{U}_x^{-1} \mathbf{A}_x - \mathbf{U}_y^{-1} \mathbf{A}_y\|_F^2 + \lambda_y \|\mathbf{U}_y^{-1}\|_F^2 \end{aligned} \quad (14)$$

450 The above problems are also ridge regression problems and
 451 have analytical solutions of \mathbf{U}_x and \mathbf{U}_y as follows:
 452

$$453 \quad \begin{aligned} \mathbf{U}_x^{-1} &= \mathbf{U}_y^{-1} \mathbf{A}_y \mathbf{A}_x^T (\mathbf{A}_x \mathbf{A}_x^T + (\gamma_x / \alpha) \mathbf{I})^{-1} \\ \mathbf{U}_y^{-1} &= \mathbf{U}_x^{-1} \mathbf{A}_x \mathbf{A}_y^T (\mathbf{A}_y \mathbf{A}_y^T + (\gamma_y / \alpha) \mathbf{I})^{-1} \end{aligned} \quad (15)$$

456 Here, we verify that \mathbf{U}_x^{-1} and \mathbf{U}_y^{-1} are invertible. The
 457 \mathbf{U}_x^{-1} and \mathbf{U}_y^{-1} are both initialized as an identity matrix, of
 458 suitable dimension, which is invertible. That is, we have
 459 $\mathbf{U}_y^{(0)} = \mathbf{I}$ when we compute \mathbf{U}_x^{-1} . If $\mathbf{A}_y \mathbf{A}_x^T$ is invertible,
 460 then \mathbf{U}_x^{-1} is invertible. In fact, we have $\mathbf{A}_y, \mathbf{A}_x \in \mathbb{R}^{d \times N}$.
 461 d is the dimension of the sample. For a patch of size 8×8 ,
 462 $d = 64$. The N is the number of samples in the training
 463 data. Remember that we have much more samples when
 464 compared to the dimension of patches, that is $N \gg d$. It is
 465 less likely that $\mathbf{A}_y \mathbf{A}_x^T \in \mathbb{R}^{d \times d}$ has a rank structure lower
 466 than d . In other words, $\mathbf{A}_y \mathbf{A}_x^T \in \mathbb{R}^{d \times d}$ is less likely to be
 467 singular if we have enough training data. The experiments
 468 also confirm our conjecture. Besides, we can also add small
 469 disburcation to guarantee that $\mathbf{A}_y \mathbf{A}_x^T \in \mathbb{R}^{d \times d}$ is invertible.
 470

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

473 **3.3. Real Image Denoising** 500

474 Two methods:

475 The first one is that

$$477 \quad \begin{aligned} & \min_{\mathbf{a}_{x,i}, \mathbf{a}_{y,i}} \|\mathbf{x}_i - \mathbf{D}_x \mathbf{a}_{x,i}\|_2^2 + \|\mathbf{y}_i - \mathbf{D}_y \mathbf{a}_{y,i} - \mathbf{v}_{y,i}\|_2^2 \\ & \quad + \alpha \|\mathbf{U}_x^{-1} \mathbf{a}_{x,i} - \mathbf{U}_y^{-1} \mathbf{a}_{y,i}\|_2^2 \\ & \quad + \beta_x \|\mathbf{a}_{x,i}\|_1 + \beta_{x2} \|\mathbf{a}_{x,i}\|_2^2 + \beta_y \|\mathbf{a}_{y,i}\|_1 + \beta_{y2} \|\mathbf{a}_{y,i}\|_2^2 \\ & \quad (+\gamma_y \|\mathbf{v}_{y,i}\|_1) \end{aligned} \quad (16)$$

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

The second one is to solve

$$486 \quad \begin{aligned} & \min_{\mathbf{a}_{y,i}, \mathbf{v}_{y,i}} \|\mathbf{y}_i - \mathbf{D}_y \mathbf{a}_{y,i} - \mathbf{v}_{y,i}\|_2^2 + \alpha \|\mathbf{U}_x^{-1} \mathbf{a}_{x,i} - \mathbf{U}_y^{-1} \mathbf{a}_{y,i}\|_2^2 \\ & \quad + \beta_{y1} \|\mathbf{a}_{y,i}\|_1 + \beta_{y2} \|\mathbf{a}_{y,i}\|_2^2 \\ & \quad (+\gamma_y \|\mathbf{v}_{y,i}\|_1). \end{aligned} \quad (17)$$

493 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 $\hat{\mathbf{x}}_i \approx$
 494 $\mathbf{D}_x \hat{\mathbf{a}}_{x,i}$.

495 Experiments demonstrate that the first method can get
 496 better performance than the second one while the second
 497 one can get faster speed than the first one.

498 We can also initialized the solution from the second one.

500 **4. The Overall Algorithm** 501502 **4.1. Pair Sample Construction from Unpaired Sam-
503 ples** 504

505 In cross style transfer methods such as CDL and SCDL,
 506 the authors assume that the two different styles have paired
 507 data, i.e., for each data sample in one style, we can find
 508 paired data sample in the other style. However, in real
 509 world, the data from two different sources may be un-
 510 paired. For example, the real noisy images should not
 511 have groundtruth clean images of the same scene. The real
 512 low-resolution images should not have corresponding high-
 513 resolution images in the real world. The real blurry images
 514 should not have corresponding clear and high quality im-
 515 ages in real world.

516 To deal with unpaired data, we could collect real noisy
 517 images and clean natural images from two different sources.
 518 The real noisy images are from the example images (18 im-
 519 ages) of the Neat Image website while the clean natural im-
 520 ages are from the training set (200 images) of the Berkeley
 521 Segmentation Dataset (BSDS500). To make use of the un-
 522 paired data samples, we employ searching strategy to con-
 523 struct the training dataset. That is, for each noisy image
 524 patch, we utilize the k-Nearest Neighbor (k-NN) algorithm
 525 to find the most similar patch in the clean images as the
 526 paired groundtruth patch. The similarity is measured by the
 527 Euclidean distance (also called squared error or ℓ_2 norm).

528 **4.2. Structual Clustering and Model Selection** 529

530 In fact, different image structures should have differ-
 531 ent influences on dictioanry as well as the mapping func-
 532 tion. Patches with flat region should have low rank struc-
 533 ture within dictionary elements and identity mapping be-
 534 tween noisy and latent clean patches. Patches with com-
 535 plex details should have more comprehensive dictionary el-
 536 ements within dictionary elements and more complex map-
 537 ping function between noisy and clean patches. A single
 538 mapping function cannot deal with all these complex rela-
 539 tionships. Hence, a structural clustering procedure is needed

540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
for complex solution. In this paper, we propose to employ Gaussian Mixture Model to cluster different image patches into different groups and learn dictionary and mapping function for each group.

4.3. Adaptive Iterations of Different Noise Levels

547
548
549
550
551
552
553
554
For real image denoising, we can perform well on images which have similar noise levels with the training dataset. How can we deal with the real noisy images whose noise levels are higher than the training dataset? The answer is to remove the noise by more iterations. The input image of each iteration is the recovered image of previous iteration. This makes sense since we can still view the recovered image as a real noisy image.

555
556
557
558
559
560
561
562
563
564
565
This will also bring a second problem, that how we could automatically terminate the iteration. This can be solved by two methods. One way is to compare the images between two iterations and calculate their difference, the iteration can be terminated if the difference is smaller than a threshold. The other way is to estimate the noise level of the current image and terminate the iterations when the noise level is lower than a preset threshold. We employ the second way and set the threshold as 0.0001 in our experiments. In fact, most of our testing images will be denoised well in one iteration.

4.4. Efficient Model Selection by Gating Network

566
567
568
569
570
571
In the Gaussian component selection procedure, if we employ the full posterior estimation, the speed is not fast. Our algorithm can be speeded up by introducing the Gating network model.

5. Experiments

572
573
574
575
576
577
578
579
580
We compare with popular software NeatImage which is one of the best denoising software available. All these methods need noise estimation which is vary hard to perform if there is no uniform regions are available in the testing image. The NeatImage will fail to perform automatical parameters settings if there is no uniform regions.

5.1. Parameters

581
582
583
584
We don't fine tune the parameters both in the training and testing datasets.

5.2. Real Image Denoising

587
588
589
590
591
592
593
We compare the proposed method with the famous BM3D [5] and WNNM [9], Cascade of Shrinkage Fields (CSF) [10], trainable reaction diffusion (TRD) [12], plain neural network based method MLP [8], the blind image denoising method Noise Clinic [17], and the commercial software Neat Image. The RGB images are firstly transformed into YCbCr channels and restored by these methods. Then

594
595
596
597
598
the denoised RGB image is obtained by transforming the restored YCbCr image back.

599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
We evaluate the competing denoising methods from various research directions on two datasets. Both the two datasets comes from the [19]. The first contains 3 cropped images of size 512×512 . The other dataset contains 42 images cropped to size of 500×500 from the 17 images provided in [19]. The 60 images contain most of the scenes in the 17 images [19].

6. Conclusion and Future Work

615
616
617
618
619
620
621
In the future, we will evaluate the proposed method on other computer vision tasks such as single image super-resolution, photo-sketch synthesis, and cross-domain image recognition. Our proposed method can be improved if we use better training images, fine tune the parameters via cross-validation. We believe that our framework can be useful not just for real image denoising, but for image super-resolution, image cross-style synthesis, and recognition tasks. This will be our line of future work.

References

- [1] 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] A. Buades, B. Coll, and J. M. Morel. A non-local algorithm for image denoising. *CVPR*, pages 60–65, 2005. 1
- [3] S. Roth and M. J. Black. Fields of Experts. *International Journal of Computer Vision*, 82(2):205–229, 2009. 1
- [4] 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
- [5] 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, 6
- [6] J. Mairal, F. Bach, J. Ponce, G. Sapiro, and A. Zisserman. Non-local sparse models for image restoration. *ICCV*, pages 2272–2279, 2009. 1
- [7] D. Zoran and Y. Weiss. From learning models of natural image patches to whole image restoration. *ICCV*, pages 479–486, 2011. 1
- [8] 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, 6
- [9] S. Gu, L. Zhang, W. Zuo, and X. Feng. Weighted nuclear norm minimization with application to image denoising. *CVPR*, pages 2862–2869, 2014. 1, 6

648 Table 1. Average PSNR(dB) results of different methods on 3 real noisy images captured by Canon EOS 5D mark3 at ISO3200 in [19]. 702
649

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

655
656 Table 2. Average SSIM results of different methods on 3 real noisy images captured by Canon EOS 5D mark3 at ISO3200 in [19]. 710
657

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

- [10] 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, 6
- [11] 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
- [12] 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, 6
- [13] J. Portilla. Full blind denoising through noise covariance estimation using gaussian scale mixtures in the wavelet domain. *Image Processing, 2004. ICIP '04. 2004 International Conference on*, 2:1217–1220, 2004. 1, 3
- [14] Tamer Rabie. Robust estimation approach for blind denoising. *Image Processing, IEEE Transactions on*, 14(11):1755–1765, 2005. 1, 3
- [15] C. Liu, R. Szeliski, S. Bing Kang, C. L. Zitnick, and W. T. Freeman. Automatic estimation and removal of noise from a single image. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(2):299–314, 2008. 1, 3
- [16] Zheng Gong, Zuowei Shen, and Kim-Chuan Toh. Image restoration with mixed or unknown noises. *Multiscale Modeling & Simulation*, 12(2):458–487, 2014. 1, 3
- [17] M. Lebrun, M. Colom, and J.-M. Morel. Multiscale image blind denoising. *Image Processing, IEEE Transactions on*, 24(10):3149–3161, 2015. 1, 3, 6
- [18] Fengyuan Zhu, Guangyong Chen, and Pheng-Ann Heng. From noise modeling to blind image denoising. *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2016. 1, 3
- [19] 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

- [20] 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
- [21] C. M. Bishop. *Pattern recognition and machine learning*. New York: Springer, 2006. 1, 3
- [22] 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. 2
- [23] 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
- [24] 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
- [25] 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
- [26] 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

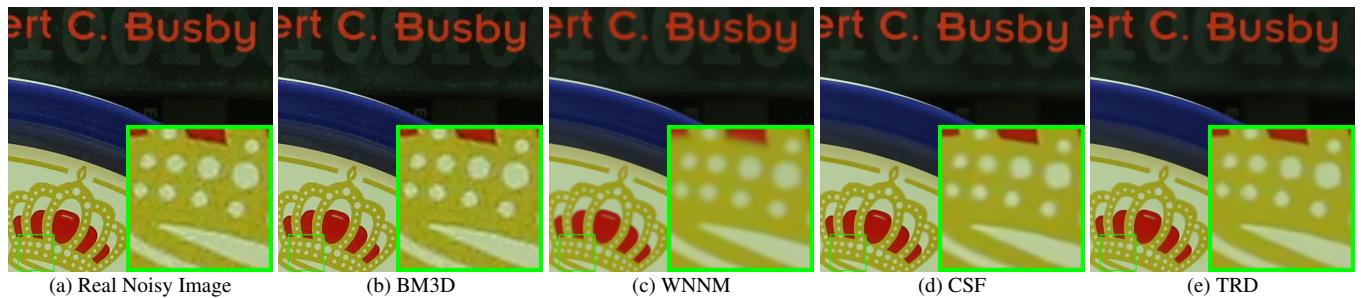
756
757
758
759
760
761
762
763
764
765
766810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863

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

777

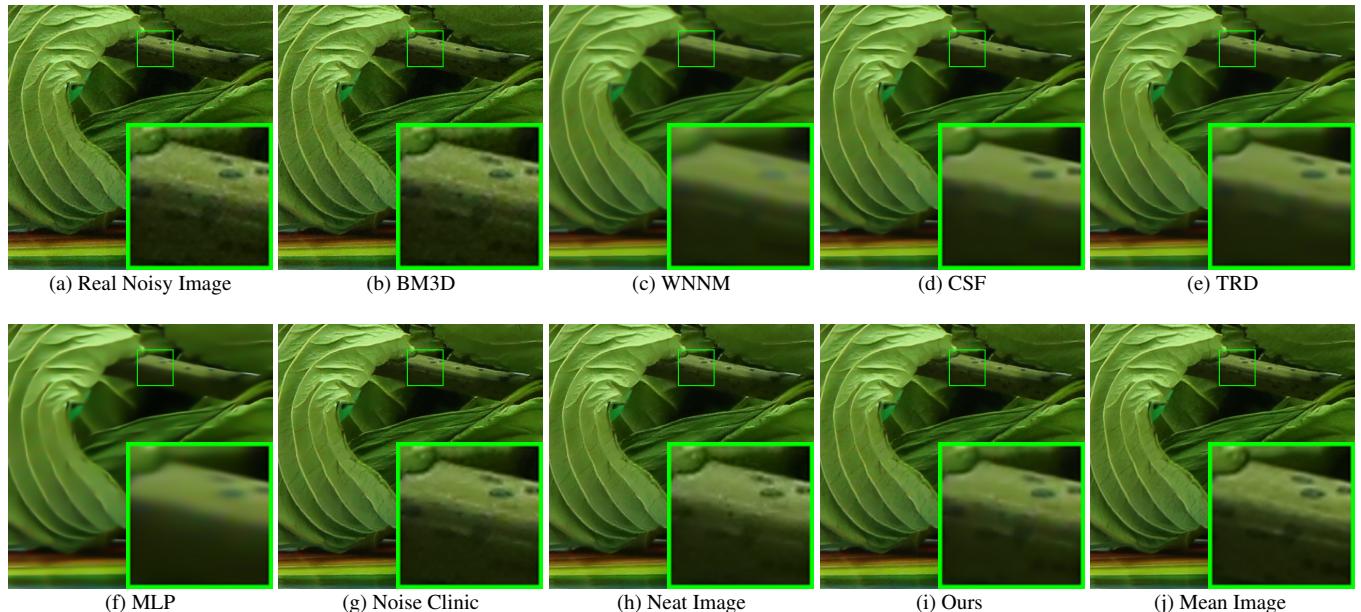


Figure 2. Denoised images of the old image "5dmark3iso32002" by different methods. The images are better to be zoomed in on screen.

801

802

803

Table 3. Average PSNR(dB) and SSIM results of different methods on 42 cropped images from 17 real noisy images in [19].

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

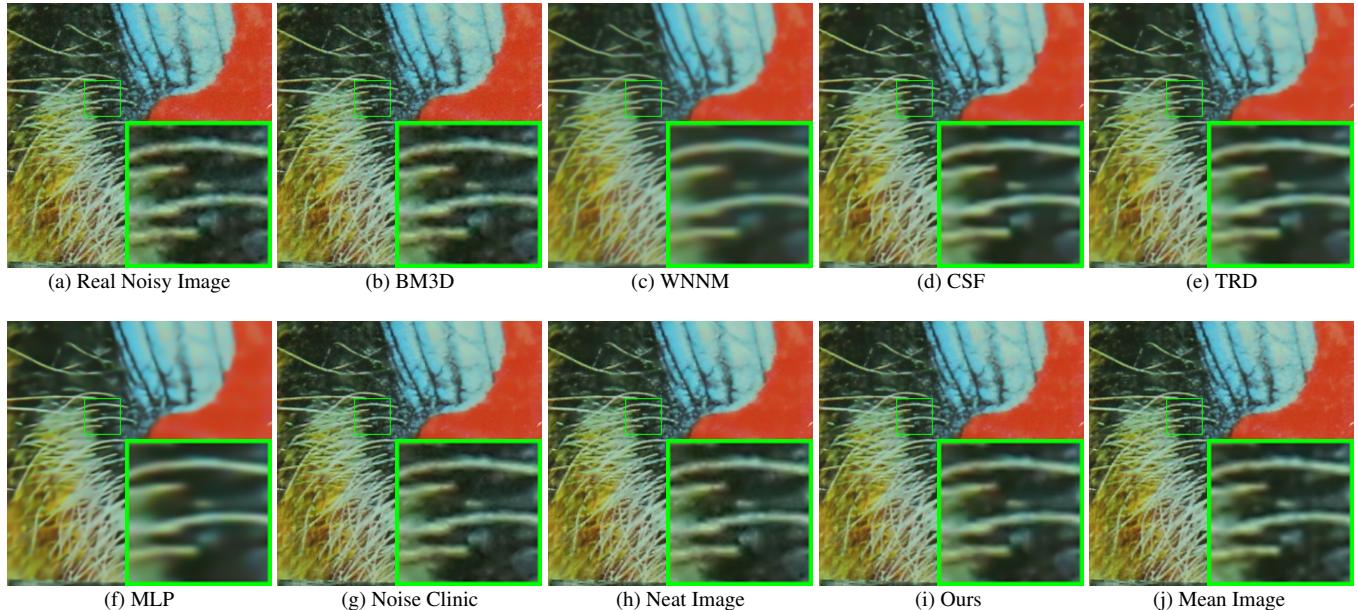
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934

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

901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971