

000  
001  
002003 

# A Noise Estimation Free Framework for Robust Real Image Denoising

004  
005  
006  
007  
008  
009  
010  
011012 Anonymous CVPR submission  
013014 Paper ID \*\*\*\*  
015016 

## Abstract

017 *Noise modeling and estimation have been two key factors  
018 for the success of traditional image denoising methods.  
019 The noise is often assumed to be homogeneous white Gaussian  
020 distributed. However, noise in real images are much  
021 more complex. Though the Mixture of Gaussians model is  
022 employed for modeling more complex noises, the estimation  
023 process is unavoidably time consuming. In this paper,  
024 we propose a novel method for real image denoising  
025 without noise modeling or estimation. Different from previous  
026 methods, we formulate the real image denoising as a cross-domain  
027 image synthesis problem. We also propose a bi-coupled dictionary learning  
028 model, in which the coefficients over dictionaries on both domains are coupled by  
029 two important regularities relationships. Experiments show  
030 that the proposed Bi-Coupled Dictionary Learning (BCDL)  
031 model is very effective in real image denoising problem, resulting  
032 competing performance on quality measurements such as visual quality, PSNR, and SSIM.*

033 

## 1. Introduction

034 Image denoising is a fundamental problem in computer  
035 vision and image processing. This problem has been studied  
036 for several decades. It is an ideal platform for testing  
037 natural image models and an important part of image  
038 enhancement for digital image acquisition systems. It also  
039 provides high-quality images and better features for other  
040 computer vision tasks such as image registration, segmentation,  
041 and pattern recognition. Over the last decades, most  
042 work focus on the removal of additive white Gaussian noise  
043 (AWGN) [1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12]. Though being  
044 effective at Gaussian noise removal, these methods would  
045 possibly fail at other types of noise, especially noise in real  
046 images, since noise in real images are much more complex  
047 than Gaussian [17]. Besides, these methods lack the  
048 ability of noise estimation and can not perform real image  
049 denoising without the help of noise estimation algorithms.  
050 In order to do real image denoising, these methods require  
051 noise estimation as a preprocessing. This make a two-stage  
052

053 framework: a thorough noise estimation method followed  
054 by an adapted denoising method. Under this framework,  
055 many representative methods achieve good performance in  
056 real image denoising [13, 14, 15, 16, 17, 18, 19].

057 The noise model in real images are commonly assumed  
058 to be Gaussian [13, 14, 15], mixed Gaussian and Lapla-  
059 cian [16], Mixture of Gaussians [17, 18], and tensor Gaus-  
060 sian across the RGB channels [19]. Although Gaussian  
061 assumption is commonly made in image denoising, it is  
062 inflexible for more complex noise in real images, as sug-  
063 gested in [15, 17]. However, as pointed out in [19], the  
064 real camera noise is far from being Gaussian, signal depen-  
065 dent, and usually hard to estimate. Although the Mixture  
066 of Gaussian model can be used to approximate any con-  
067 tinuous distribution [22], estimating its parameters is time  
068 consuming through the commonly employed Expectation-  
069 Maximization (EM) algorithm [21] or variational inference  
070 technique [22]. On the other hand, the method [16] needs  
071 tune specific parameters for different types of noise. This  
072 makes the proposed method not a strictly real image de-  
073 noising. Hence, it is still desirable to design an robust and  
074 effective model for real image denoising, which bases on  
075 few assumptions and can deal with unknown noise with-  
076 out any parameter tuning procedure. The in-camera imag-  
077 ing pipeline includes image demosaicing, white balance and  
078 color space transform, gamut mapping, tone mapping, and  
079 JPEG compression. Finally, the major noise in the real im-  
080 age can be categorized into five different types: fixed pat-  
081 tern noise, dark current noise, short noise, amplifier noise,  
082 and quantization noise [20].

083 In this paper, we avoid the challenge problem of noise es-  
084 timation. We therefore propose a cross domain synthesis so-  
085 lution for real image denoising. In fact, we propose a novel  
086 double semi-coupled dictionary learning algorithm for real  
087 image denoising problem. In the training stage, given the  
088 training patches (noisy ones and clean ones), the dictionaries  
089 and coefficients of both the clean and noisy patches. The  
090 mapping between the clean and noisy coefficients matrices  
091 are also learned in our model. Given a real noisy image,  
092 we first extract overlapping patches from it. Then we obtain  
093 the clean coefficient from an optimization framework sim-  
094

108  
 109  
 110  
 111  
 112  
 113  
 114  
 115  
 116  
 117  
 118  
 119  
 120  
 121  
 122  
 123  
 124  
 125  
 126  
 127  
 128  
 129  
 130  
 131  
 132  
 133  
 134  
 135  
 136  
 137  
 138  
 139  
 140  
 141  
 142  
 143  
 144  
 145  
 146  
 147  
 148  
 149  
 150  
 151  
 152  
 153  
 154  
 155  
 156  
 157  
 158  
 159  
 160  
 161

ilar to the training stage. The recovered patches are reconstructed by the obtained coefficients on corresponding clean dictionary atoms. We perform comprehensive experiments on real noisy images from multiple different CMOS or CCS sensors. The results demonstrate that our method achieves comparable or even better denoising performance (PSNR, SSIM, and visual quality) on most real noisy images. This reveals that the proposed method has the substantial effect of cross domain image synthesis framework for real image denoising task.

## 1.1. Our Contributions

To summarize, the contributions of this paper are as follows:

- 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. They later imposed relaxed constraints on the observed dictionary/coefficient pairs across image domains for improved performance. Wang et al. [24] further proposed a semi-coupled dictionary learning (SCDL) scheme by advancing a linear mapping for cross-domain image sparse representation. Their method has been successfully applied to applications of image SR and cross-style synthesis.

## 2.2. Real Image Denoising

To the best of our knowledge, the study of real image denoising can be dated back to the BLS-GSM model [25], in which Portilla et al. proposed to use scale mixture of Gaussian in overcomplete oriented pyramids to estimate the latent clean images. In [13], Portilla proposed to use a correlated Gaussian model for noise estimation of each wavelet subband. Based on the robust statistics theory [?], the work of Rabie [14] modeled the noisy pixels as outliers, which could be removed via Lorentzian robust estimator. In [15], Liu et al. proposed to use ‘noise level function’ (NLF) to estimate the noise and then use Gaussian conditional random field to obtain the latent clean image. Recently, Gong et al. proposed an optimization based method [16], which models the data fitting term by weighted sum of  $\ell_1$  and  $\ell_2$  norms and the regularization term by sparsity prior in the wavelet transform domain. Later, Lebrun et al. proposed a multiscale denoising algorithm called ‘Noise Clinic’ [17] for real image denoising task. This method generalizes the NL-Bayes [26] to deal with signal, scale, and frequency dependent noise. Recently, Zhu et al. proposed a Bayesian model [18] which approximates the noise via Mixture of Gaussian (MoG) model [22]. The clean image is recovered from the noisy image by the proposed Low Rank MoG filter (LR-MoG). However, noise level estimation is already a challenging problem and denoising methods are quite sensitive to this parameter. Moreover, these methods are based on shrinkage models that are too simple to reflect reality, which results in over-smoothing of important structures such as small-scale text and textures.

## 3. Double Semi-Couple Dictionary Learning

In this section, we first formulate the real image denoising problem from the perspective of learning based model and then provide the optimization for the problem.

### 3.1. Problem Formulation

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

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

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,  $\ell_1$  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 images and latent clean natural images.

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

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  $\leq 1$  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 varible 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.

### 3.2.1 Updating $\mathbf{A}_x$ and $\mathbf{A}_y$

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

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

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

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

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

326

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

327

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

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

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

341

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

342

343

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

344

345 These two are quadraically constrained quadratic program (QCQP) problem and can be solved by Lagrange dual techniques.

### 3.2.3 Updating $\mathbf{V}_y$

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

356

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

357

358 This is a ridge regression problem. We can obtain the analytical solution of  $\mathbf{V}_y$  by

359

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

360

### 3.2.4 Alternate Updating $\mathbf{V}_y$

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

367

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

368

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

370

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

371

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

372

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

373

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

375

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

376

377 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 invertible. 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 invertible, then  $\mathbf{U}_x^{-1}$  is invertible. 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  $8 \times 8$ ,  $d = 64$ . 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 disbturtion to guarantee that  $\mathbf{A}_y \mathbf{A}_x^T \in \mathbb{R}^{d \times d}$  is invertible.

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

## 3.3 Real Image Denoising

406 Two methods:

407 The first one is that

408

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

409

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

411 The second one is to solve

412

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

413

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

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

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

432     **4. The Overall Algorithm**  
433  
434     **4.1. Pair Sample Construction from Unpaired Sam-**  
435     **ples**

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

To deal with unpaired data, we could collect real noisy images and clean natural images from two different sources. The real noisy images are from the example images (18 images) of the Neat Image website while the clean natural images are from the training set (200 images) of the Berkeley Segmentation Dataset (BSDS500). To make use of the unpaired data samples, we employ searching strategy to construct the training dataset. That is, for each noisy image patch, we utilize the k-Nearest Neighbor (k-NN) algorithm to find the most similar patch in the clean images as the paired groundtruth patch. The similarity is measured by the Euclidean distance (also called squared error or  $\ell_2$  norm).

460     **4.2. Structural Clustering and Model Selection**  
461

In fact, different image structures should have different influences on dictionary as well as the mapping function. Patches with flat region should have low rank structure within dictionary elements and identity mapping between noisy and latent clean patches. Patches with complex details should have more comprehensive dictionary elements within dictionary elements and more complex mapping function between noisy and clean patches. A single mapping function cannot deal with all these complex relationships. Hence, a structural clustering procedure is needed 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.

476     **4.3. Adaptive Iterations of Different Noise Levels**  
477

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.

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.

498     **4.4. Efficient Model Selection by Gating Network**  
499

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.

500     **5. Experiments**  
501

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.

514     **5.1. Parameters**  
515

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

516     **5.2. Real Image Denoising**  
517

We compare the proposed method with the famous BM3D [5] and WNNM [9], Cascade of Shrinkage Fields (CSF) [10], trainable reaction diffusion (TRD) [11], plain neural network based method MLP [12], 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 the denoised RGB image is obtained by transforming the restored YCbCr image back.

531     **5.3. Comparison on Dataset with GroundTruth**  
532

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  $512 \times 512$ . The other dataset contains 42 images cropped to size of  $500 \times 500$  from the 17 images provided in [19]. The 60 images contain most of the scenes in the 17 images [19].



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

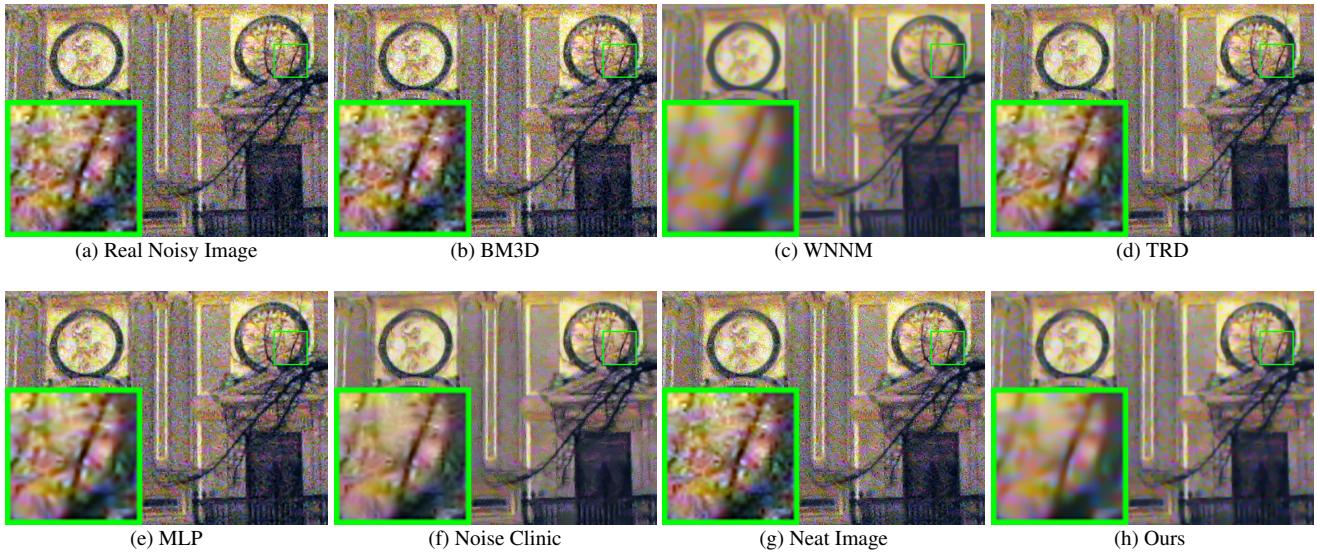


Figure 2. Denoised images of the old image "Palace" by different methods. The images are better to be zoomed in on screen.

## 6. Conclusion and Future Work

In the future, we will evaluate the proposed method on other computer vision tasks such as single image super-

resolution, photo-sketch synthesis, and cross-domain image recognition. Our proposed method can be improved if we use better training images, fine tune the parameters

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		703
650		704
651		705
652		706
653		707
654		708
655		709
656		710
657		711
658		712
659		713
660		714
661		715
662		716
663		717
664		718
665		719
666		720
667		721
668		722
669		723
670		724
671		725
672		726
673		727
674		728
675		729
676		730
677		731
678		732
679		733
680		734
681		735
682		736
683		737
684		738
685		739
686		740
687		741
688		742
689		743
690		744
691		745
692		746
693		747
694		748
695		749
696		750
697		751
698		752
699		753
700		754
701		755

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	<b>38.76</b>	37.68	38.63
2	33.88	33.95	33.95	34.90	35.04	31.94	35.69	34.87	<b>35.96</b>
3	33.83	33.85	33.85	34.15	34.07	30.89	<b>35.54</b>	34.77	35.51
Average	34.90	34.96	34.96	35.50	35.54	31.91	36.67	35.77	<b>36.70</b>

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

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	<b>0.9712</b>
2	0.8919	0.8848	0.8951	0.9159	0.9187	0.8498	0.9427	0.9308	<b>0.9434</b>
3	0.9128	0.9136	0.9136	0.9254	0.9279	0.8635	0.9476	0.9463	<b>0.9529</b>
Average	0.9131	0.9117	0.9153	0.9337	0.9358	0.8725	0.9531	0.9457	<b>0.9558</b>

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] 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](#)
- [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, 5](#)
- [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] 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](#)
- [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, 5](#)
- [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, 5](#)
- [11] Yunjin Chen, Wei Yu, and Thomas Pock. On learning optimized reaction diffusion processes for effective image restoration. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 5261–5269, 2015. [1, 5](#)
- [12] 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](#)
- [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, 2](#)
- [14] Tamer Rabie. Robust estimation approach for blind denoising. *Image Processing, IEEE Transactions on*, 14(11):1755–1765, 2005. [1, 2](#)
- [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, 2](#)
- [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, 2](#)
- [17] 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](#)
- [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, 2](#)
- [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, 5, 7, 9](#)

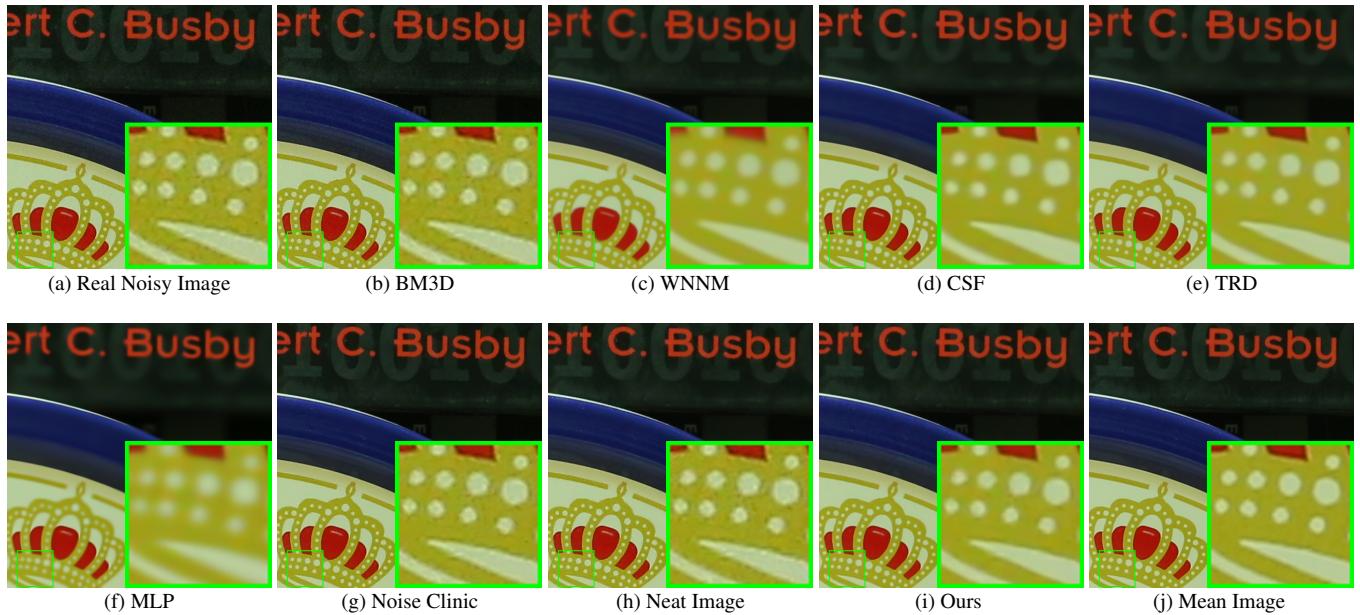


Figure 3. Denoised images of the old image "5dmark3iso3200<sub>1</sub>" by different methods. The images are better to be zoomed in on screen.

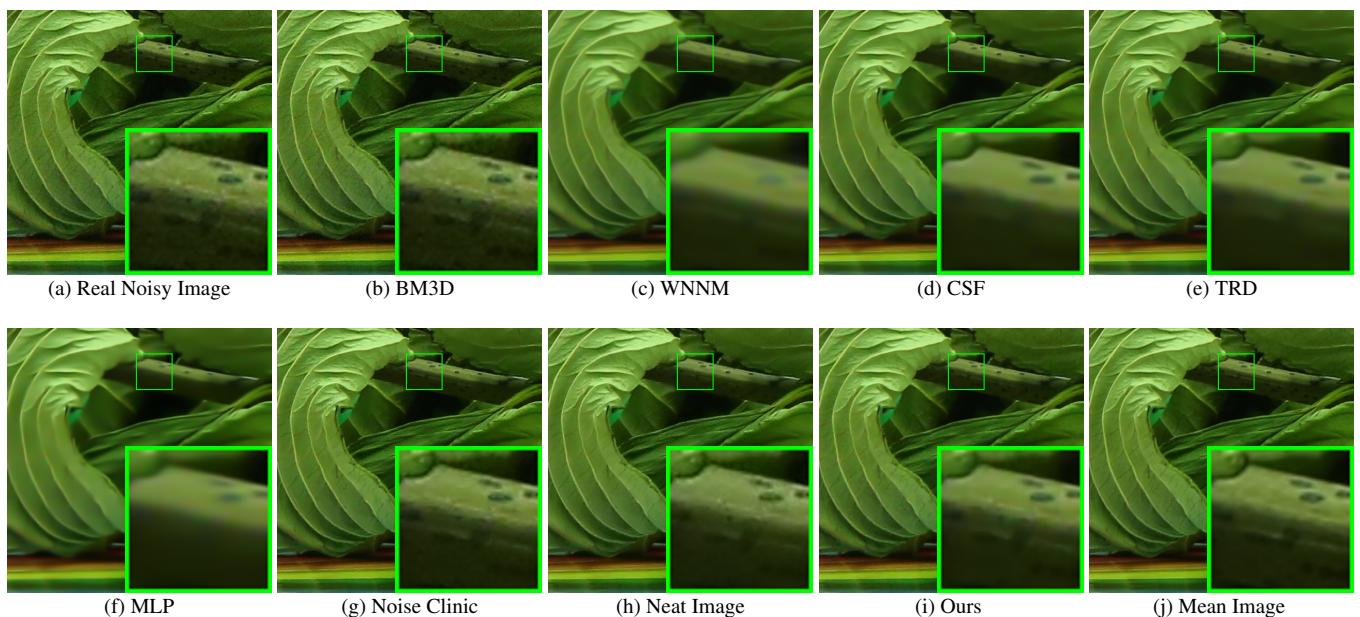


Figure 4. Denoised images of the old image "5dmark3iso3200<sub>2</sub>" by different methods. The images are better to be zoomed in on screen.

- [20] 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
- [21] 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
- [22] C. M. Bishop. *Pattern recognition and machine learning*. New York: Springer, 2006. 1, 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.

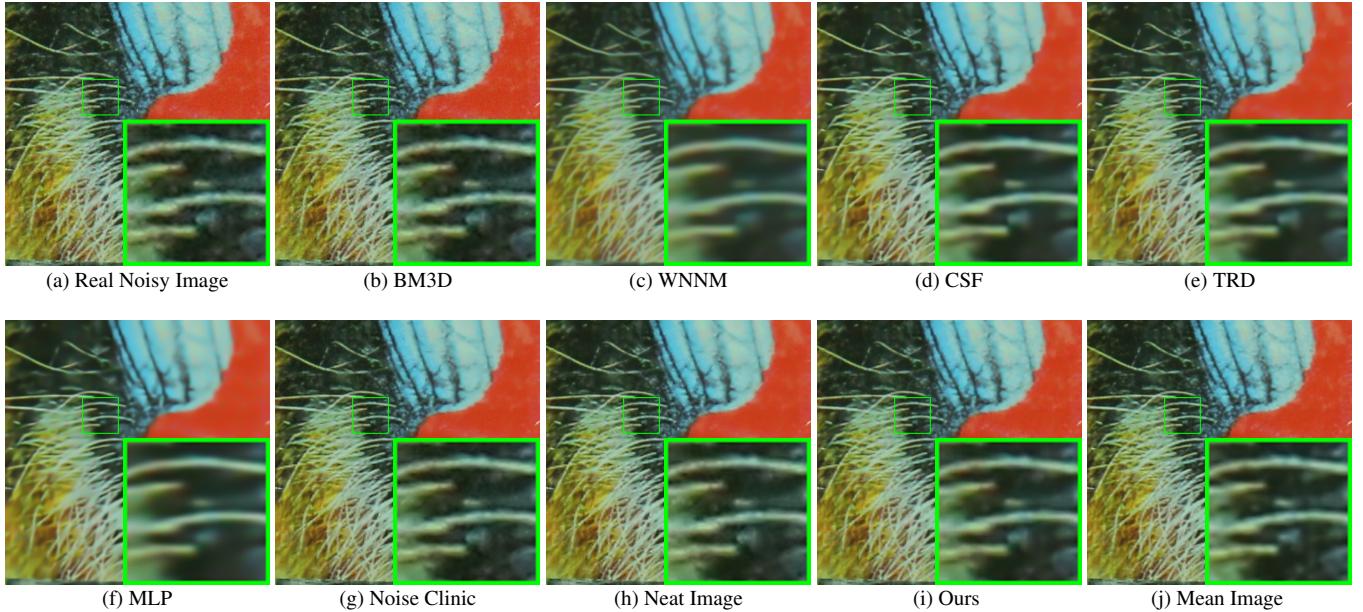


Figure 5. Denoised images of the old image "5dmark3iso32003" by different methods. The images are better to be zoomed in on screen.

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

892 Image denoising using scale mixtures of Gaussians in the  
893 wavelet domain. *Image Processing, IEEE Transactions on*,  
894 12(11):1338–1351, 2003. 2

- 895 [26] M. Lebrun, A. Buades, and J. M. Morel. A nonlocal bayesian  
896 image denoising algorithm. *SIAM Journal on Imaging Sciences*,  
897 6(3):1665–1688, 2013. 2

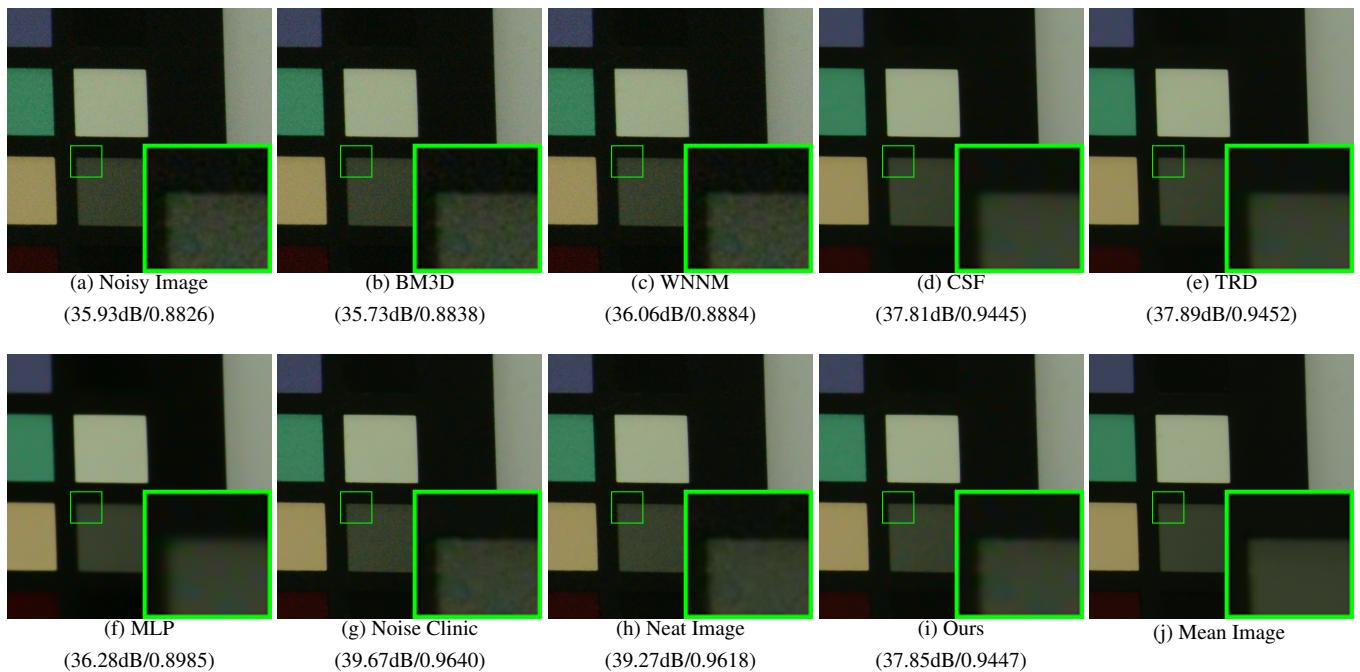


Figure 6. Denoised images of the old image "NikonD600ISO3200C37" by different methods. The images are better to be zoomed in on screen.

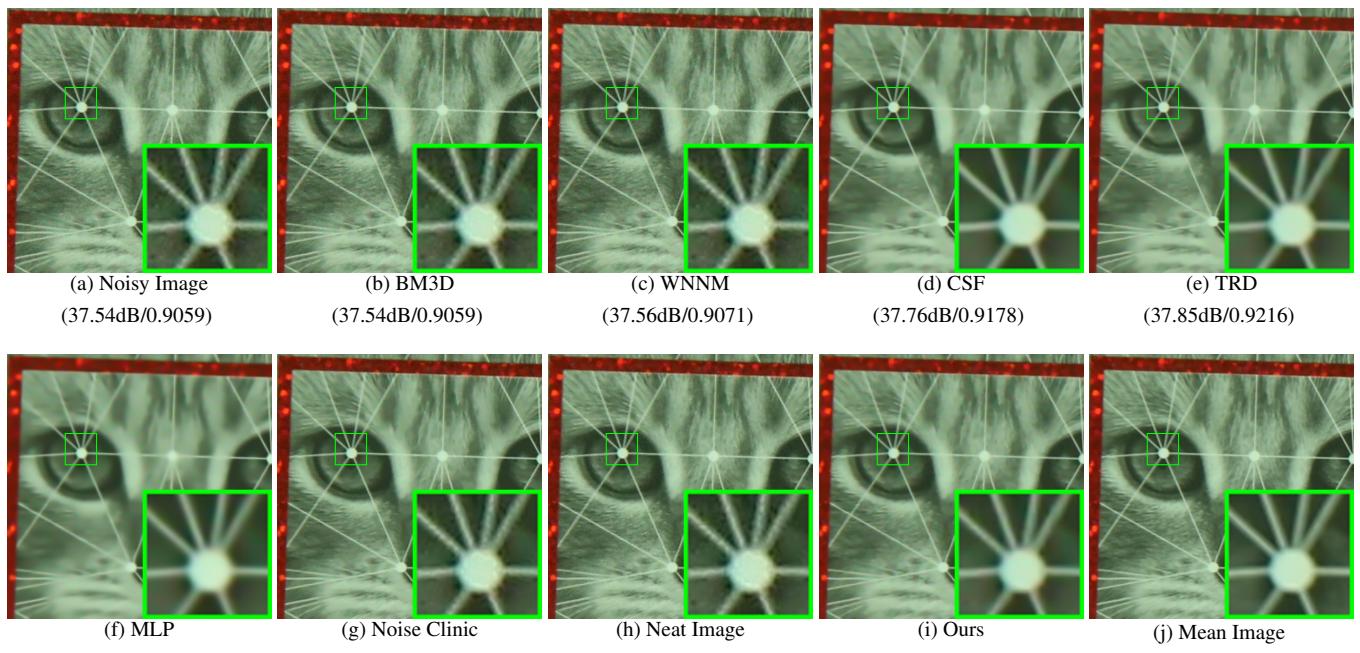


Figure 7. Denoised images of the old image "CanonEOS5DMark3ISO3200C326" by different methods. The images are better to be zoomed in on screen.

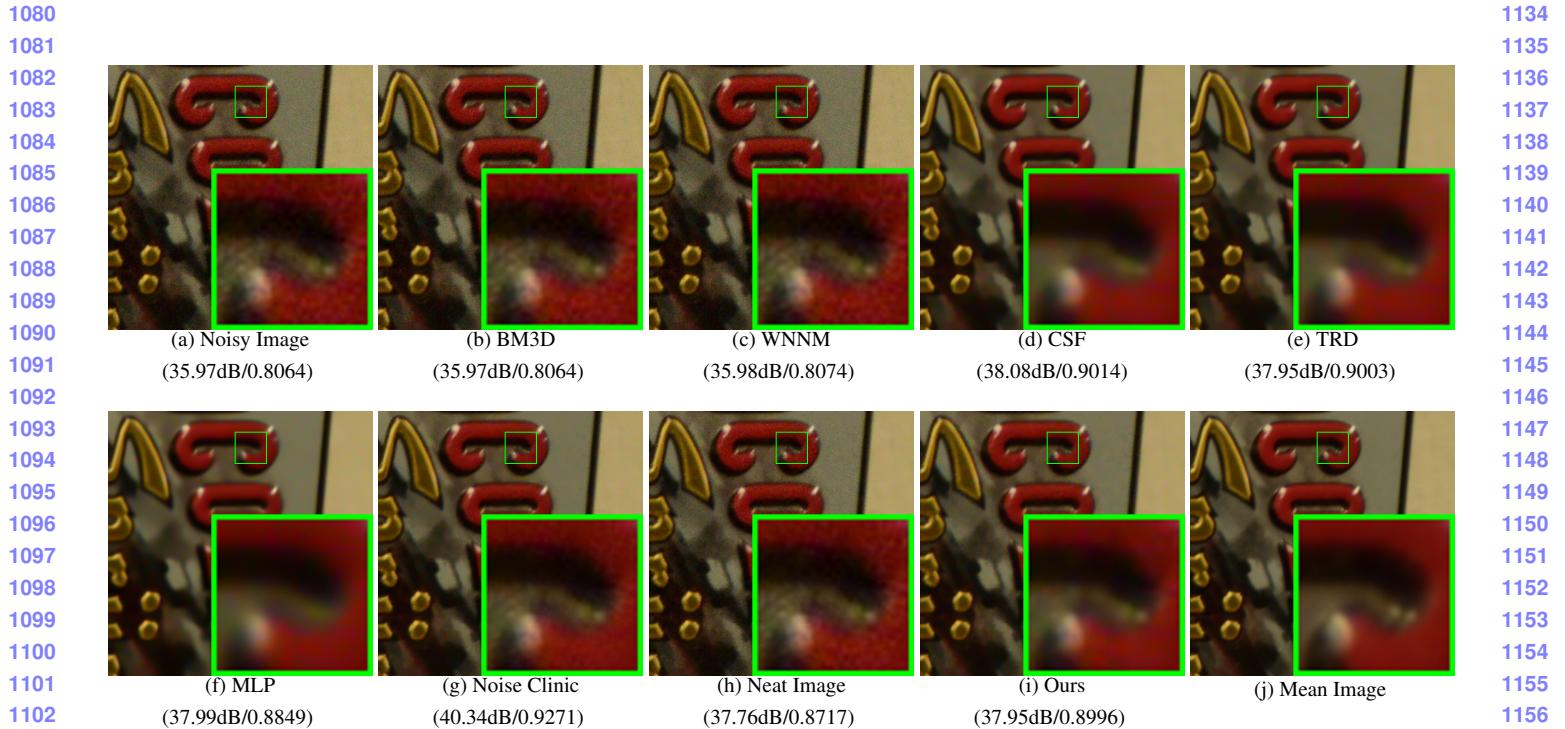


Figure 8. Denoised images of the old image "NikonD800ISO3200A113" by different methods. The images are better to be zoomed in on screen.

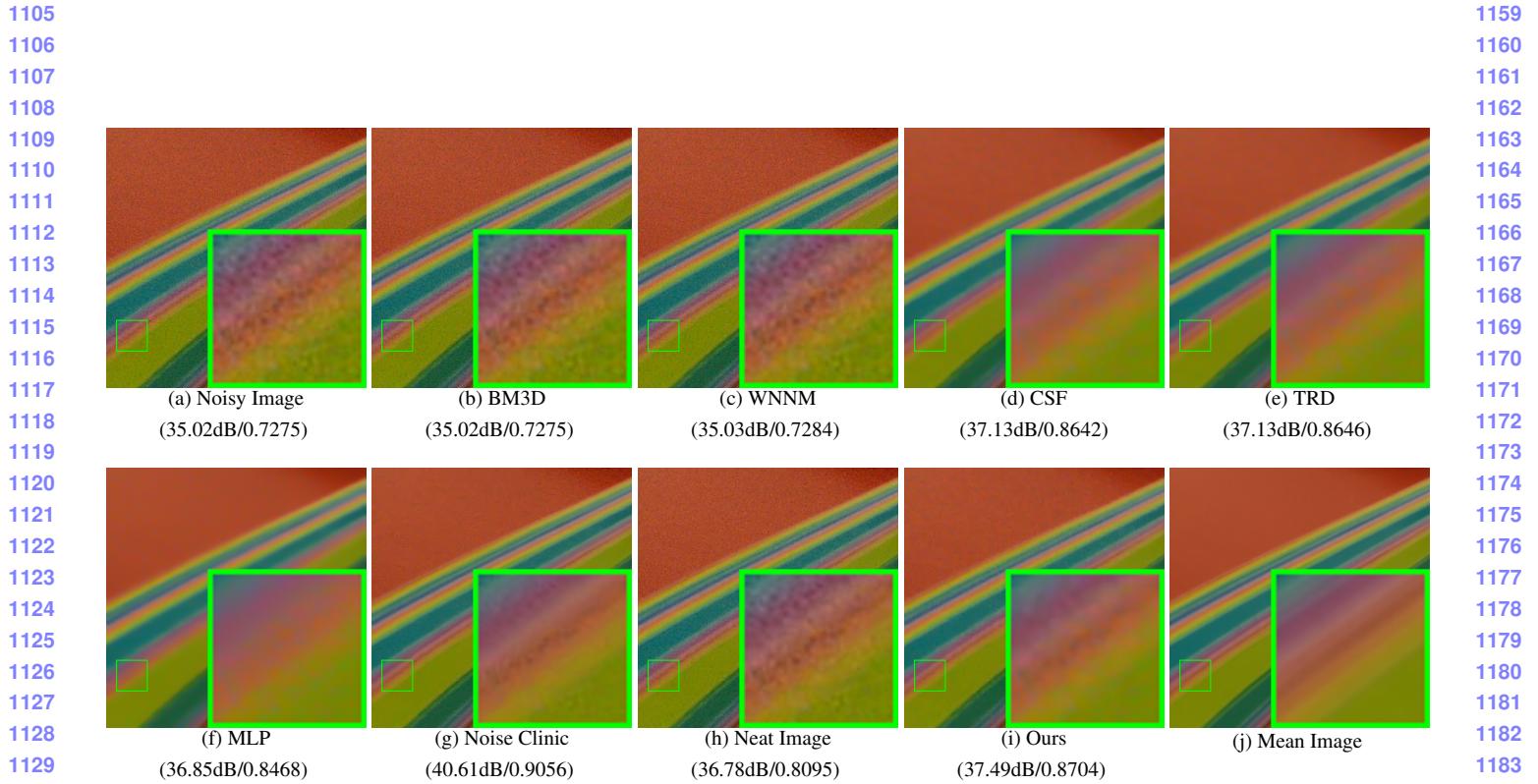


Figure 9. Denoised images of the old image "NikonD800ISO3200A1130" by different methods. The images are better to be zoomed in on screen.

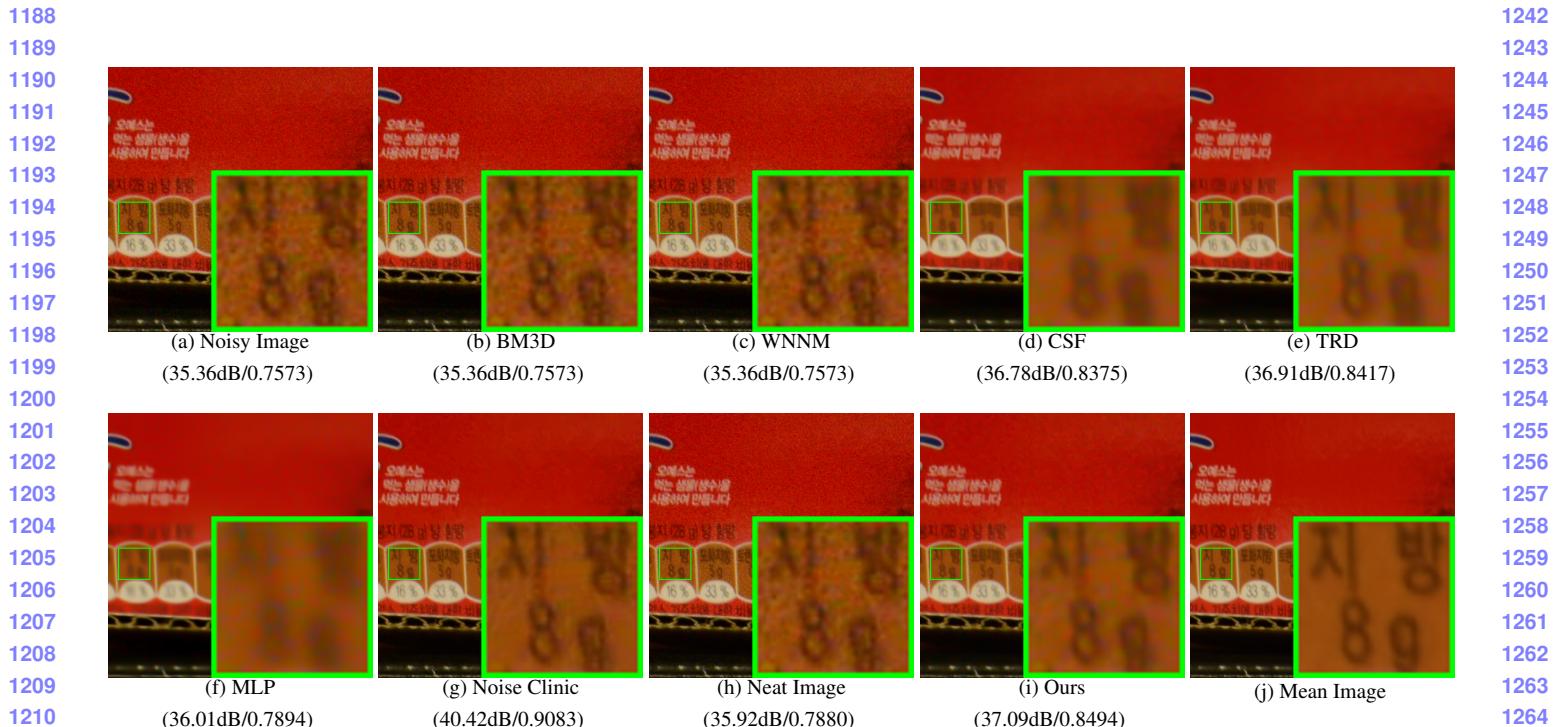


Figure 10. Denoised images of the old image "Nikon\_D800\_ISO3200\_A1\_21" by different methods. The images are better to be zoomed in on screen.

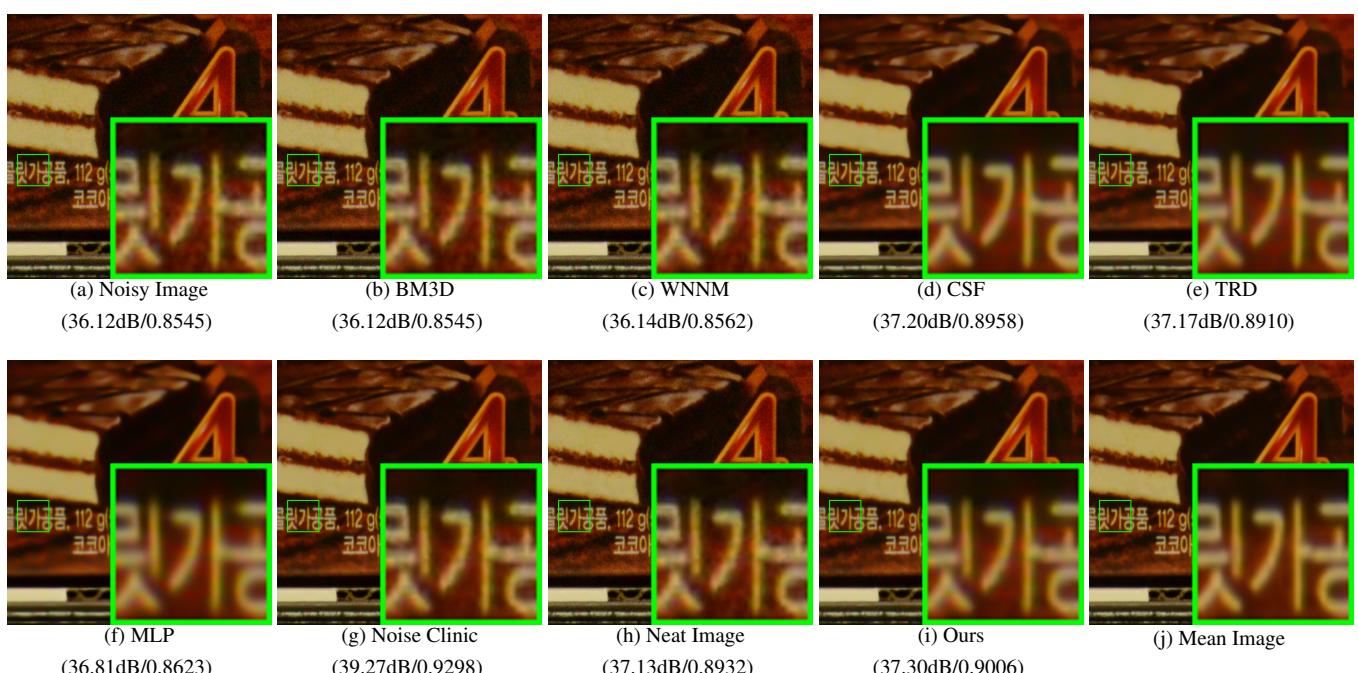


Figure 11. Denoised images of the old image "Nikon\_D800\_ISO3200\_A1\_23" by different methods. The images are better to be zoomed in on screen.

1296

1297

1298

1299

1300

1301

1302

1303

1304

1305

1306

1307

1308

1309

1310

1311

1312

1313

1314

1315

1316

1317

1318

1319

1320

1321

1322

1323

1324

1325

1326

1327

1328

1329

1330

1331

1332

1333

1334

1335

1336

1337

1338

1339

1340

1341

1342

1343

1344

1345

1346

1347

1348

1349

1350

1351

1352

1353

1354

1355

1356

1357

1358

1359

1360

1361

1362

1363

1364

1365

1366

1367

1368

1369

1370

1371

1372

1373

1374

1375

1376

1377

1378

1379

1380

1381

1382

1383

1384

1385

1386

1387

1388

1389

1390

1391

1392

1393

1394

1395

1396

1397

1398

1399

1400

1401

1402

1403



Figure 12. Denoised images of the old image "Nikon\_D800\_ISO3200\_A299" by different methods. The images are better to be zoomed in on screen.

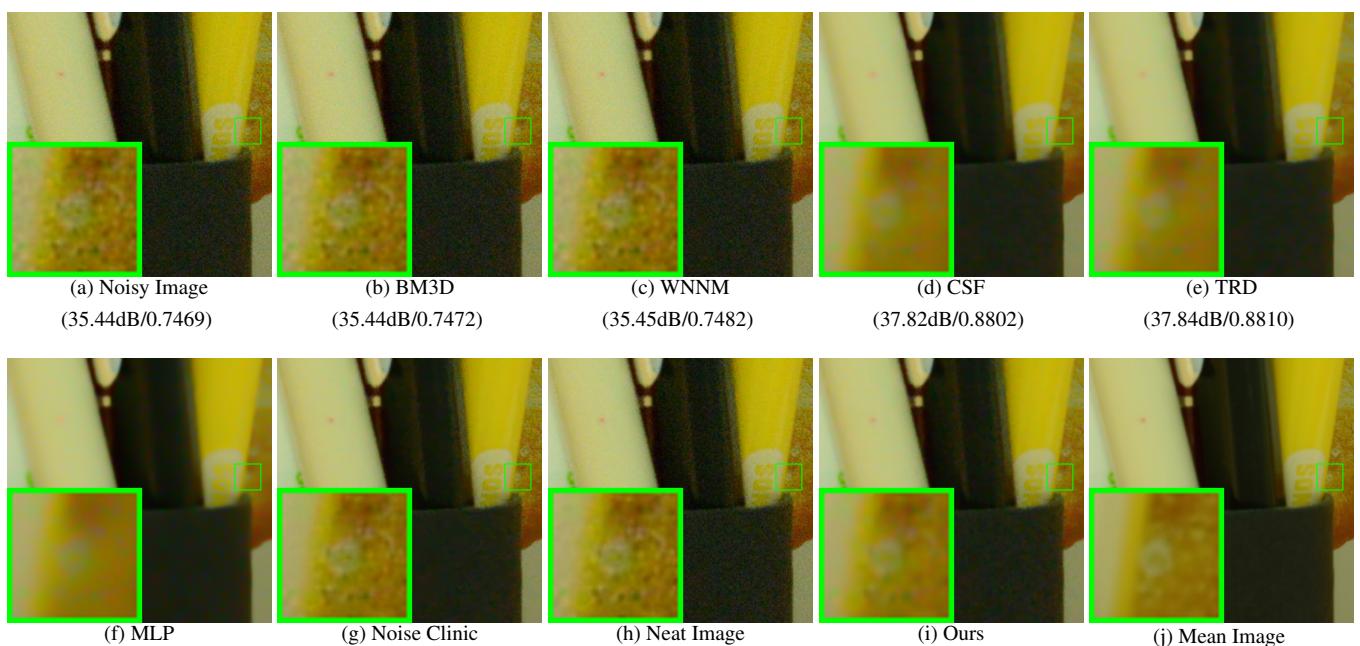


Figure 13. Denoised images of the old image "Nikon\_D800\_ISO6400\_B2\_13" by different methods. The images are better to be zoomed in on screen.