Toward Convolutional Blind Denoising of Real Photographs

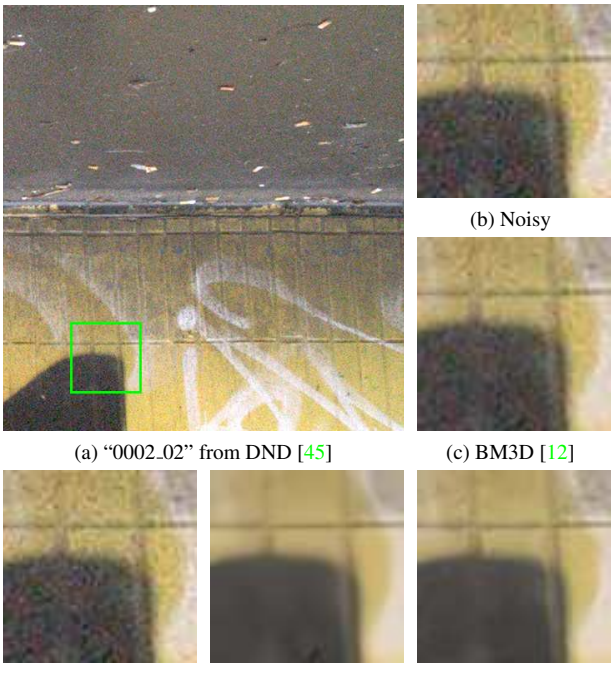
Shi Guo[1,3,4], Zifei Yan(✉) [1], Kai Zhang[1,3], Wangmeng Zuo[1,2], Lei Zhang[3,4]
[1]Harbin Institute of Technology, Harbin; [2]Peng Cheng Laboratory, Shenzhen;
[3] The Hong Kong Polytechnic University, Hong Kong; [4]DAMO Academy, Alibaba Group
guoshi28@outlook.com, {wmzuo,yanzifei}@hit.edu.cn
cskaizhang@gmail.com, cslzhang@comp.polyu.edu.hk

Abstract

While deep convolutional neural networks (CNNs) have achieved impressive success in image denoising with additive white Gaussian noise (AWGN), their performance remains limited on real-world noisy photographs. The main reason is that their learned models are easy to overfit on the simplified AWGN model which deviates severely from the complicated real-world noise model. In order to improve the generalization ability of deep CNN denoisers, we suggest training a convolutional blind denoising network (CBDNet) with more realistic noise model and real-world noisy-clean image pairs. On the one hand, both signal-dependent noise and in-camera signal processing pipeline is considered to synthesize realistic noisy images. On the other hand, real-world noisy photographs and their nearly noise-free counterparts are also included to train our CBDNet. To further provide an interactive strategy to rectify denoising result conveniently, a noise estimation subnetwork with asymmetric learning to suppress under-estimation of noise level is embedded into CBDNet. Extensive experimental results on three datasets of real-world noisy photographs clearly demonstrate the superior performance of CBDNet over state-of-the-arts in terms of quantitative metrics and visual quality. The code has been made available at https://github.com/GuoShi28/CBDNet.

1. Introduction

Image denoising is an essential and fundamental problem in low-level vision and image processing. With decades of studies, numerous promising approaches [3, 12, 17, 53, 11, 61] have been developed and near-optimal performance [8, 31, 50] has been achieved for the removal of additive white Gaussian noise (AWGN). However, in real camera system, image noise comes from multiple sources (*e.g.*, dark current noise, short noise, and thermal noise) and is further affected by in-camera processing (ISP) pipeline (*e.g.*, demosaicing, Gamma correction, and compression). All these make real noise much more different from AWGN, and blind denoising of real-world noisy photographs remains a challenging issue.



(a) "0002_02" from DND [45] (b) Noisy (c) BM3D [12] (d) DnCNN [61] (e) FFDNet+ [62] (f) **CBDNet**

Figure 1: Denoising results of different methods on real-world noisy image "0002_02" from DND [45].

In the recent past, Gaussian denoising performance has been significantly advanced by the development of deep CNNs [61, 38, 62]. However, deep denoisers for blind AWGN removal degrades dramatically when applied to real photographs (see Fig. 1(d)). On the other hand, deep denoisers for non-blind AWGN removal would smooth out the details while removing the noise (see Fig. 1(e)). Such an phenomenon may be explained from the characteristic of deep CNNs [39], where their generalization largely depends on the ability of memorizing large scale training data. In other words, existing CNN denoisers tend to be over-fitted to Gaussian noise and generalize poorly to real-world noisy images with more sophisticated noise.

In this paper, we tackle this issue by developing a convolutional blind denoising network (CBDNet) for real-world photographs. As indicated by [39], the success of CNN denoisers are significantly dependent on whether the distributions of synthetic and real noises are well matched. Therefore, realistic noise model is the foremost issue for blind denoising of real photographs. According to [14, 45], Poisson-Gaussian distribution which can be approximated as heteroscedastic Gaussian of a signal-dependent and a stationary noise components has been considered as a more appropriate alternative than AWGN for real raw noise modeling. Moreover, in-camera processing would further makes the noise spatially and chromatically correlated which increases the complexity of noise. As such, we take into account both Poisson-Gaussian model and in-camera processing pipeline (*e.g.*, demosaicing, Gamma correction, and JPEG compression) in our noise model. Experiments show that in-camera processing pipeline plays a pivot role in realistic noise modeling, and achieves notably performance gain (*i.e.*, > 5 dB by PSNR) over AWGN on DND [45].

We further incorporate both synthetic and real noisy images to train CBDNet. On one hand, it is easy to access massive synthetic noisy images. However, the noise in real photographs cannot be fully characterized by our model, thereby giving some leeway for improving denoising performance. On the other hand, several approaches [43, 1] have suggested to get noise-free image by averaging hundreds of noisy images at the same scene. Such solution, however, is expensive in cost, and suffers from the over-smoothing effect of noise-free image. Benefited from the incorporation of synthetic and real noisy images, $0.3 \sim 0.5$ dB gain on PSNR can be attained by CBDNet on DND [45].

Our CBDNet is comprised of two subnetworks, *i.e.*, noise estimation and non-blind denoising. With the introduction of noise estimation subnetwork, we adopt an asymmetric loss by imposing more penalty on under-estimation error of noise level, making our CBDNet perform robustly when the noise model is not well matched with real-world noise. Besides, it also allows the user to interactively rectify the denoising result by tuning the estimated noise level map. Extensive experiments are conducted on three real noisy image datasets, *i.e.*, NC12 [29], DND [45] and Nam [43]. In terms of both quantitative metrics and perceptual quality, our CBDNet performs favorably in comparison to state-of-the-arts. As shown in Fig. 1, both non-blind BM3D [12] and DnCNN for blind AWGN [61] fail to denoise the real-world noisy photograph. In contrast, our CBDNet achieves very pleasing denoising results by retaining most structure and details while removing the sophisticated real-world noise.

To sum up, the contribution of this work is four-fold:

- A realistic noise model is presented by considering both heteroscedastic Gaussian noise and in-camera processing pipeline, greatly benefiting the denoising performance.
- Synthetic noisy images and real noisy photographs are incorporated for better characterizing real-world image noise and improving denoising performance.
- Benefited from the introduction of noise estimation subnetwork, asymmetric loss is suggested to improve the generalization ability to real noise, and interactive denoising is allowed by adjusting the noise level map.
- Experiments on three real-world noisy image datasets show that our CBDNet achieves state-of-the-art results in terms of both quantitative metrics and visual quality.

2. Related Work

2.1. Deep CNN Denoisers

The advent of deep neural networks (DNNs) has led to great improvement on Gaussian denoising. Until Burger *et al*. [6], most early deep models cannot achieve state-of-the-art denoising performance [22, 49, 57]. Subsequently, CSF [53] and TNRD [11] unroll the optimization algorithms for solving the fields of experts model to learn stage-wise inference procedure. By incorporating residual learning [19] and batch normalization [21], Zhang *et al*. [61] suggest a denoising CNN (DnCNN) which can outperform traditional non-CNN based methods. Without using clean data, Noise2Noise [30] also achieves state-of-the-art. Most recently, other CNN methods, such as RED30 [38], MemNet [55], BM3D-Net [60], MWCNN [33] and FFDNet [62], are also developed with promising denoising performance.

Benefited from the modeling capability of CNNs, the studies [61, 38, 55] show that it is feasible to learn a single model for blind Gaussian denoising. However, these blind models may be over-fitted to AWGN and fail to handle real noise. In contrast, non-blind CNN denoisiers, *e.g.*, FFDNet [62], can achieve satisfying results on most real noisy images by manually setting proper or relatively higher noise level. To exploit this characteristic, our CBDNet includes a noise estimation subnetwork as well as an asymmetric loss to suppress under-estimation error of noise level.

2.2. Image Noise Modeling

Most denoising methods are developed for non-blind Gaussian denoising. However, the noise in real images comes from various sources (dark current noise, short noise, thermal noise, etc.), and is much more sophisticated [44]. By modeling photon sensing with Poisson and remaining stationary disturbances with Gaussian, Poisson-Gaussian noise model [14] has been adopted for the raw data of imaging sensors. In [14, 32], camera response function (CRF) and quantization noise are also considered for more practical noise modeling. Instead of Poisson-Gaussian, Hwang *et*

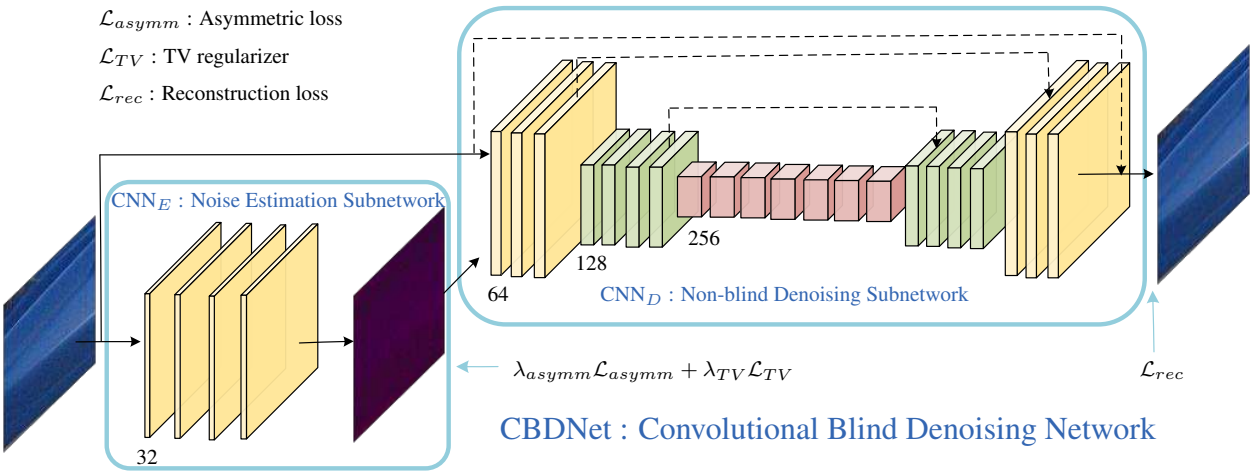


Figure 2: Illustration of our CBDNet for blind denoising of real-world noisy photograph.

al. [20] present a Skellam distribution for Poisson photon noise modeling. Moreover, when taking in-camera image processing pipeline into account, the channel-independent noise assumption may not hold true, and several approaches [25, 43] are proposed for cross-channel noise modeling. In this work, we show that realistic noise model plays a pivot role in CNN-based denoising of real photographs, and both Poisson-Gaussian noise and in-camera image processing pipeline benefit denoising performance.

2.3. Blind Denoising of Real Images

Blind denoising of real noisy images generally is more challenging and can involve two stages, *i.e*., noise estimation and non-blind denoising. For AWGN, several PCA-based [48, 34, 9] methods have been developed for estimating noise standard deviation ($SD.$). Rabie [49] models the noisy pixels as outliers and exploits Lorentzian robust estimator for AWGN estimation. For Poisson-Gaussian model, Foi *et al*. [14] suggest a two-stage scheme, *i.e*., local estimation of multiple expectation/standard-deviation pairs, and global parametric model fitting.

In most blind denoising methods, noise estimation is closely coupled with non-blind denoising. Portilla [46, 47] adopts a Gaussian scale mixture for modeling wavelet patches of each scale, and utilizes Bayesian least square to estimate clean wavelet patches. Based on the piecewise smooth image model, Liu *et al*. [32] propose a unified framework for the estimation and removal of color noise. Gong *et al*. [15] model the data fitting term as the weighted sum of the L_1 and L_2 norms, and utilize a sparsity regularizer in wavelet domain for handling mixed or unknown noises. Lebrun *et al*. [28, 29] propose an extension of non-local Bayes approach [27] by modeling the noise of each patch group to be zero-mean correlated Gaussian distributed. Zhu *et al*. [63] suggest a Bayesian nonparametric technique to remove the noise via the low-rank mixture of Gaussians (LR-MoG) model. Nam *et al*. [43] model the cross-channel noise as a multivariate Gaussian and perform denoising by the Bayesian nonlocal means filter [24]. Xu *et al*. [59] suggest a multi-channel weighted nuclear norm minimization (MCWNNM) model to exploit channel redundancy. They further present a trilateral weighted sparse coding (TWSC) method for better modeling noise and image priors [58]. Except noise clinic (NC) [28, 29], MCWNNM [59], and TWSC [58], the codes of most blind denoisers are not available. Our experiments show that they are still limited for removing noise from real images.

3. Proposed Method

This section presents our CBDNet consisting of a noise estimation subnetwork and a non-blind denoising subnetwork. To begin with, we introduce the noise model to generate synthetic noisy images. Then, the network architecture and asymmetric loss. Finally, we explain the incorporation of synthetic and real noisy images for training CBDNet.

3.1. Realistic Noise Model

As noted in [39], the generalization of CNN largely depends on the ability in memorizing training data. Existing CNN denoisers, *e.g*., DnCNN [61], generally does not work well on real noisy images, mainly due to that they may be over-fitted to AWGN while the real noise distribution is much different from Gaussian. On the other hand, when trained with a realistic noise model, the memorization ability of CNN will be helpful to make the learned model generalize well to real photographs. Thus, noise model plays a critical role in guaranteeing performance of CNN denoiser.

Different from AWGN, real image noise generally is more sophisticated and signal-dependent [35, 14]. Practically, the noise produced by photon sensing can be modeled as Poisson, while the remaining stationary disturbances

can be modeled as Gaussian. Poisson-Gaussian thus provides a reasonable noise model for the raw data of imaging sensors [14], and can be further approximated with a heteroscedastic Gaussian $\mathbf{n}(\mathbf{L}) \sim \mathcal{N}(0, \sigma^2(\mathbf{L}))$ defined as,

$$\sigma^2(\mathbf{L}) = \mathbf{L} \cdot \sigma_s^2 + \sigma_c^2. \quad (1)$$

where $\mathbf{L}$ is the irradiance image of raw pixels. $\mathbf{n}(\mathbf{L}) = \mathbf{n}_s(\mathbf{L}) + \mathbf{n}_c$ involves two components, *i.e.*, a stationary noise component $\mathbf{n}_c$ with noise variance σ_c^2 and a signal-dependent noise component $\mathbf{n}_s$ with spatially variant noise variance $\mathbf{L} \cdot \sigma_s^2$.

Real photographs, however, are usually obtained after in-camera processing (ISP), which further increases the complexity of noise and makes it spatially and chromatically correlated. Thus, we take two main steps of ISP pipeline, *i.e.*, demosaicing and Gamma correction, into consideration, resulting in the realistic noise model as,

$$\mathbf{y} = f(\mathbf{DM}(\mathbf{L} + \mathbf{n}(\mathbf{L}))), \quad (2)$$

where $\mathbf{y}$ denotes the synthetic noisy image, $f(\cdot)$ stands for the camera response function (CRF) uniformly sampled from the 201 CRFs provided in [16]. And $\mathbf{L} = \mathbf{M}f^{-1}(\mathbf{x})$ is adopted to generate irradiance image from a clean image $\mathbf{x}$. $\mathbf{M}(\cdot)$ represents the function that converts sRGB image to Bayer image and $\mathbf{DM}(\cdot)$ represents the demosaicing function [37]. Note that the interpolation in $\mathbf{DM}(\cdot)$ involves pixels of different channels and spatial locations. The synthetic noise in Eqn. (2) is thus channel and space dependent.

Furthermore, to extend CBDNet for handling compressed image, we can include JPEG compression in generating synthetic noisy image,

$$\mathbf{y} = JPEG(f(\mathbf{DM}(\mathbf{L} + \mathbf{n}(\mathbf{L})))). \quad (3)$$

For noisy uncompressed image, we adopt the model in Eqn. (2) to generate synthetic noisy images. For noisy compressed image, we exploit the model in Eqn. (3). Specifically, σ_s and σ_c are uniformly sampled from the ranges of $[0, 0.16]$ and $[0, 0.06]$, respectively. In JPEG compression, the quality factor is sampled from the range $[60, 100]$. We note that the quantization noise is not considered because it is minimal and can be ignored without any obvious effect on denoising result [62].

3.2. Network Architecture

As illustrated in Fig. 2, the proposed CBDNet includes a noise estimation subnetwork CNN_E and a non-blind denosing subnetwork CNN_D. First, CNN_E takes a noisy observation $\mathbf{y}$ to produce the estimated noise level map $\hat{\sigma}(\mathbf{y}) = \mathcal{F}_E(\mathbf{y}; \mathbf{W}_E)$, where $\mathbf{W}_E$ denotes the network parameters of CNN_E. We let the output of CNN_E be the noise level map due to that it is of the same size with the input $\mathbf{y}$ and can be estimated with a fully convolutional network. Then, CNN_D takes both $\mathbf{y}$ and $\hat{\sigma}(\mathbf{y})$ as input to obtain the final denoising result $\hat{\mathbf{x}} = \mathcal{F}_D(\mathbf{y}, \hat{\sigma}(\mathbf{y}); \mathbf{W}_D)$, where $\mathbf{W}_D$ denotes the network parameters of CNN_D. Moreover, the introduction of CNN_E also allows us to adjust the estimated noise level map $\hat{\sigma}(\mathbf{y})$ before putting it to the the non-blind denosing subnetwork CNN_D. In this work, we present a simple strategy by letting $\hat{\varrho}(\mathbf{y}) = \gamma \cdot \hat{\sigma}(\mathbf{y})$ for interactive denoising.

We further explain the network structures of CNN_E and CNN_D. CNN_E adopts a plain five-layer fully convolutional network without pooling and batch normalization operations. In each convolution (Conv) layer, the number of feature channels is set as 32, and the filter size is 3×3. The ReLU nonlinearity [42] is deployed after each Conv layer. As for CNN_D, we adopt an U-Net [51] architecture which takes both $\mathbf{y}$ and $\hat{\sigma}(\mathbf{y})$ as input to give a prediction $\hat{\mathbf{x}}$ of the noise-free clean image. Following [61], the residual learning is adopted by first learning the residual mapping $\mathcal{R}(\mathbf{y}, \hat{\sigma}(\mathbf{y}); \mathbf{W}_D)$ and then predicting $\hat{\mathbf{x}} = \mathbf{y} + \mathcal{R}(\mathbf{y}, \hat{\sigma}(\mathbf{y}); \mathbf{W}_D)$. The 16-layer U-Net architecture of CNN_E is also given in Fig. 2, where symmetric skip connections, strided convolutions and transpose convolutions are introduced for exploiting multi-scale information as well as enlarging receptive field. All the filter size is 3×3, and the ReLU nonlinearity [42] is applied after every Conv layer except the last one. Moreover, we empirically find that batch normalization helps little for the noise removal of real photographs, partially due to that the real noise distribution is fundamentally different from Gaussian.

Finally, we note that it is also possible to train a single blind CNN denoiser by learning a direct mapping from noisy observation to clean image. However, as noted in [62, 41], taking both noisy image and noise level map as input is helpful in generalizing the learned model to images beyond the noise model and thus benefits blind denoising. We empirically find that single blind CNN denoiser performs on par with CBDNet for images with lower noise level, and is inferior to CBDNet for images with heavy noise. Furthermore, the introduction of noise estimation subnetwork also makes interactive denoising and asymmetric learning allowable. Therefore, we suggest to include the noise estimation subnetwork in our CBDNet.

3.3. Asymmetric Loss and Model Objective

Both CNN and traditional non-blind denoisers perform robustly when the input noise $SD.$ is higher than the ground-truth one (*i.e.*, over-estimation error), which encourages us to adopt asymmetric loss for improving generalization ability of CBDNet. As illustrates in FFDNet [62], BM3D/FFDNet achieve the best result when the input noise $SD.$ and ground-truth noise $SD.$ are matched. When the input noise $SD.$ is lower than the ground-truth one, the results of BM3D/FFDNet contain perceptible noises. When the input noise $SD.$ is higher than the ground-truth one, BM3D/FFDNet can still achieve satisfying results by gradually wiping out some low contrast structure along with

the increase of input noise SD. Thus, non-blind denoisers are sensitive to under-estimation error of noise SD., but are robust to over-estimation error. With such property, BM3D/FFDNnet can be used to denoise real photographs by setting relatively higher input noise SD., and this might explain the reasonable performance of BM3D on the DND benchmark [45] in the non-blind setting.

To exploit the asymmetric sensitivity in blind denoising, we present an asymmetric loss on noise estimation to avoid the occurrence of under-estimation error on the noise level map. Given the estimated noise level $\hat{\sigma}(y_i)$ at pixel i and the ground-truth $\sigma(y_i)$, more penalty should be imposed to their MSE when $\hat{\sigma}(y_i) < \sigma(y_i)$. Thus, we define the asymmetric loss on the noise estimation subnetwork as,

$$\mathcal{L}_{asymm} = \sum_i |\alpha - \mathbb{I}_{(\hat{\sigma}(y_i)-\sigma(y_i))<0}| \cdot (\hat{\sigma}(y_i) - \sigma(y_i))^2 \ , \quad (4)$$

where $\mathbb{I}_e = 1$ for $e < 0$ and 0 otherwise. By setting $0 < \alpha < 0.5$, we can impose more penalty to under-estimation error to make the model generalize well to real noise.

Furthermore, we introduce a total variation (TV) regularizer to constrain the smoothness of $\hat{\sigma}(\mathbf{y})$,

$$\mathcal{L}_{TV} = \|\nabla_h \hat{\sigma}(\mathbf{y})\|_2^2 + \|\nabla_v \hat{\sigma}(\mathbf{y})\|_2^2 , \quad (5)$$

where ∇_h (∇_v) denotes the gradient operator along the horizontal (vertical) direction. For the output $\hat{\mathbf{x}}$ of non-blind denoising, we define the reconstruction loss as,

$$\mathcal{L}_{rec} = \|\hat{\mathbf{x}} - \mathbf{x}\|_2^2 . \quad (6)$$

To sum up, the overall objective of our CBDNet is,

$$\mathcal{L} = \mathcal{L}_{rec} + \lambda_{asymm}\mathcal{L}_{asymm} + \lambda_{TV}\mathcal{L}_{TV} , \quad (7)$$

where λ_{asymm} and λ_{TV} denote the tradeoff parameters for the asymmetric loss and TV regularizer, respectively. In our experiments, the PSNR/SSIM results of CBDNet are reported by minimizing the above objective. As for qualitative evaluation of visual quality, we train CBDNet by further adding perceptual loss [23] on `relu3_3` of VGG-16 [54] to the objective in Eqn. (7).

3.4. Training with Synthetic and Real Noisy Images

The noise model in Sec. 3.1 can be used to synthesize any amount of noisy images. And we can also guarantee the high quality of the clean images. Even though, the noise in real photographs cannot be fully characterized by the noise model. Fortunately, according to [43, 45, 1], nearly noise-free image can be obtained by averaging hundreds of noisy images from the same scene, and several datasets have been built in literatures. In this case, the scenes are constrained to be static, and it is generally expensive to acquire hundreds of noisy images. Moreover, the nearly noise-free image tends to be over-smoothing due to the averaging effect. Therefore, synthetic and real noisy images can be combined to improve the generalization ability to real photographs.

In this work, we use the noise model in Sec. 3.1 to generate the synthetic noisy images, and use 400 images from BSD500 [40], 1600 images from Waterloo [36], and 1600 images from MIT-Adobe FiveK dataset [7] as the training data. Specifically, we use the RGB image $\mathbf{x}$ to synthesize clean raw image $\mathbf{L} = \mathbf{M}f^{-1}(\mathbf{x})$ as a reverse ISP process and use the same f to generate noisy image as Eqns. (2) or (3), where f is a CRF randomly sampled from those in [16]. As for real noisy images, we utilize the 120 images from the RENOIR dataset [4]. In particular, we alternatingly use the batches of synthetic and real noisy images during training. For a batch of synthetic images, all the losses in Eqn. (7) are minimized to update CBDNet. For a batch of real images, due to the unavailability of ground-truth noise level map, only $\mathcal{L}_{rec}$ and $\mathcal{L}_{TV}$ are considered in training. We empirically find that such training scheme is effective in improving the visual quality for denoising real photographs.

4. Experimental Results

4.1. Test Datasets

Three datasets of real-world noisy images, *i.e.*, NC12 [29], DND [45] and Nam [43], are adopted:

NC12 includes 12 noisy images. The ground-truth clean images are unavailable, and we only report the denoising results for qualitative evaluation.

DND contains 50 pairs of real noisy images and the corresponding nearly noise-free images. Analogous to [4], the nearly noise-free images are obtained by carefully post-processing of the low-ISO images. PSNR/SSIM results are obtained through the online submission system.

Nam contains 11 static scenes and for each scene the nearly noise-free image is the mean image of 500 JPEG noisy images. We crop these images into 512×512 patches and randomly select 25 patches for evaluation.

4.2. Implementation Details

The model parameters in Eqn. (7) are given by $\alpha = 0.3$, $\lambda_1 = 0.5$, and $\lambda_2 = 0.05$. Note that the noisy images from Nam [43] are JPEG compressed, while the noisy images from DND [45] are uncompressed. Thus we adopt the noise model in Eqn. (2) to train CBDNet for DND and NC12, and the model in Eqn. (3) to train CBDNet(JPEG) for Nam.

To train our CBDNet, we adopt the ADAM [26] algorithm with β_1 = 0.9. The method in [18] is adopted for model initialization. The size of mini-batch is 32 and the size of each patch is 128×128. All the models are trained with 40 epochs, where the learning rate for the first 20 epochs is 10^{-3}, and then the learning rate 5×10^{-4} is used to further fine-tune the model. It takes about three days to train our CBDNet with the MatConvNet package [56] on a Nvidia GeForce GTX 1080 Ti GPU.

Table 1: The quantitative results on the DND benchmark.

Method	Blind/Non-blind	Denoising on	PSNR	SSIM
CDnCNN-B [61]	Blind	sRGB	32.43	0.7900
EPLL [64]	Non-blind	sRGB	33.51	0.8244
TNRD [11]	Non-blind	sRGB	33.65	0.8306
NCSR [13]	Non-blind	sRGB	34.05	0.8351
MLP [6]	Non-blind	sRGB	34.23	0.8331
FFDNet [62]	Non-blind	sRGB	34.40	0.8474
BM3D [12]	Non-blind	sRGB	34.51	0.8507
FoE [52]	Non-blind	sRGB	34.62	0.8845
WNNM [17]	Non-blind	sRGB	34.67	0.8646
GCBD [10]	Blind	sRGB	35.58	0.9217
CIMM [5]	Non-blind	sRGB	36.04	0.9136
KSVD [3]	Non-blind	sRGB	36.49	0.8978
MCWNNM [59].	Blind	sRGB	37.38	0.9294
TWSC [58]	Blind	sRGB	37.94	0.9403
CBDNet(Syn)	Blind	sRGB	37.57	0.9360
CBDNet(Real)	Blind	sRGB	37.72	0.9408
CBDNet(All)	Blind	sRGB	**38.06**	**0.9421**

4.3. Comparison with State-of-the-arts

We consider four blind denoising approaches, *i.e.*, NC [29, 28], NI [2], MCWNNM [59] and TWSC [58] in our comparison. NI [2] is a commercial software and has been included into Photoshop and Corel PaintShop. Besides, we also include a blind Gaussian denoising method (*i.e.*, CDnCNN-B [61]), and three non-blind denoising methods (*i.e.*, CBM3D [12], WNNM [17], FFDNet [62]). When apply non-blind denoiser to real photographs, we exploit [9] to estimate the noise SD..

NC12. Fig. 3 shows the results of an NC12 images. All the competing methods are limited in removing noise in the dark region. In comparison, CBDNet performs favorably in removing noise while preserving salient image structures.

DND. Table 1 lists the PSNR/SSIM results released on the DND benchmark website. Undoubtedly, CDnCNN-B [61] cannot be generalized to real noisy photographs and performs very poorly. Although the noise SD. is provided, non-blind Gaussian denoisers, *e.g.*, WNNM [17], BM3D [12] and FoE [52], only achieve limited performance, mainly due to that the real noise is much different from AWGN. MCWNNM [59] and TWSC [58] are specially designed for blind denoising of real photographs, and also achieve promising results. Benefited from the realistic noise model and incorporation with real noisy images, our CBDNet achieves the highest PSNR/SSIM results, and slightly better than MCWNNM [59] and TWSC [58]. CBDNet also significantly outperforms another CNN-based denoiser, *i.e.*, CIMM [5]. As for running time, CBDNet takes about 0.4s to process an 512×512 image. Fig. 4 provides the denoising results of an DND image. BM3D and CDnCNN-B fail to remove most noise from real photograph, NC, NI, MCWNNM and TWSC still cannot remove all noise, and NI also suffers from the over-smoothing effect. In comparison, our CBDNet performs favorably in balancing noise removal and structure preservation.

Table 2: The quantitative results on the Nam dataset [43].

Method	Blind/Non-blind	PSNR	SSIM
NI [2]	Blind	31.52	0.9466
CDnCNN-B [61]	Blind	37.49	0.9272
TWSC [58]	Blind	37.52	0.9292
MCWNNM [59]	Blind	37.91	0.9322
BM3D [12]	Non-blind	39.84	0.9657
NC [29]	Blind	40.41	0.9731
WNNM [17]	Non-blind	41.04	0.9768
CBDNet	Blind	40.02	0.9687
CBDNet(JPEG)	Blind	**41.31**	**0.9784**

Table 3: PSNR/SSIM results by different noise models.

Method	DND [45]	Nam [43]
CBDNet(G)	32.52 / 0.79	37.62 / 0.9290
CBDNet(HG)	33.70 / 0.9084	38.40 / 0.9453
CBDNet(G+ISP)	37.41 / 0.9353	39.03 / 0.9563
CBDNet(HG+ISP)	**37.57 / 0.9360**	39.20 / 0.9579
CBDNet(JPEG)	—	**40.51 / 0.9745**

Nam. The quantitative and qualitative results are given in Table 2 and Fig. 5. CBDNet(JPEG) performs much better than CBDNet (*i.e.*, $\sim$ 1.3 dB by PSNR) and achieves the best performance in comparison to state-of-the-arts.

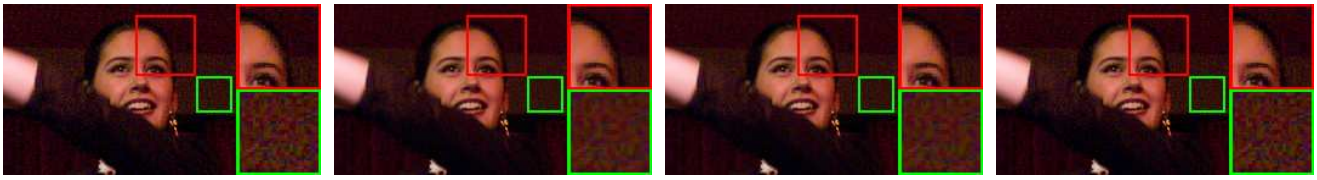
4.4. Ablation Studies

Effect of noise model. Instead of AWGN, we consider heterogeneous Gaussian (HG) and in-camera processing (ISP) pipeline for modeling image noise. On DND and Nam, we implement four variants of noise models: (i) Gaussian noise (CBDNet(G)), (ii) heterogeneous Gaussian (CBDNet(HG)), (iii) Gaussian noise and ISP (CBDNet(G+ISP)), and (iv) heterogeneous Gaussian and ISP (CBDNet(HG+ISP), *i.e.*, full CBDNet. For Nam, CBDNet(JPEG) is also included. Table 3 shows the PSNR/SSIM results of different noise models.

G vs HG. Without ISP, CBDNet(HG) achieves about $0.8 \sim 1$ dB gain over CBDNet(G). When ISP is included, the gain by HG is moderate, *i.e.*, CBDNet(HG+ISP) only outperforms CBDNet(G+ISP) about 0.15 dB.

w/o ISP. In comparison, ISP is observed to be more critical for modeling real image noise. In particular, CBDNet(G+ISP) outperforms CBDNet(G) by 4.88 dB, while CBDNet(HG+ISP) outperforms CBDNet(HG) by 3.87 dB on DND. For Nam, the inclusion of JPEG compression in ISP further brings a gain of 1.31 dB.

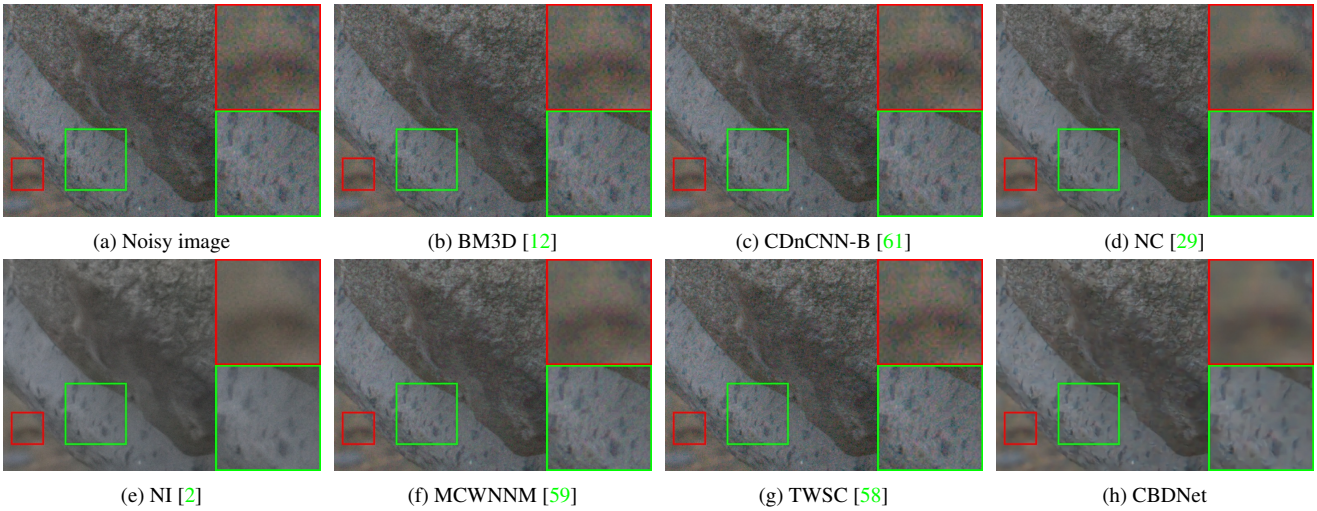
Incorporation of synthetic and real images. We implement two baselines: (i) CBDNet(Syn) trained only on synthetic images, and (ii) CBDNet(Real) trained only on real images, and rename our full CBDNet as CBDNet(All). Fig. 7 shows the denoising results of these three methods on a NC12 image. Even trained on large scale synthetic image dataset, CBDNet(Syn) still cannot remove all real noise, partially due to that real noise cannot be fully characterized by the noise model. CBDNet(Real) may produce over-smoothing results, partially due to the effect of imperfect



(a) Noisy image (b) WNNM [17] (c) FFDNet [62] (d) NC [29]

(e) NI [2] (f) MCWNNM [59] (g) TWSC [58] (h) CBDNet

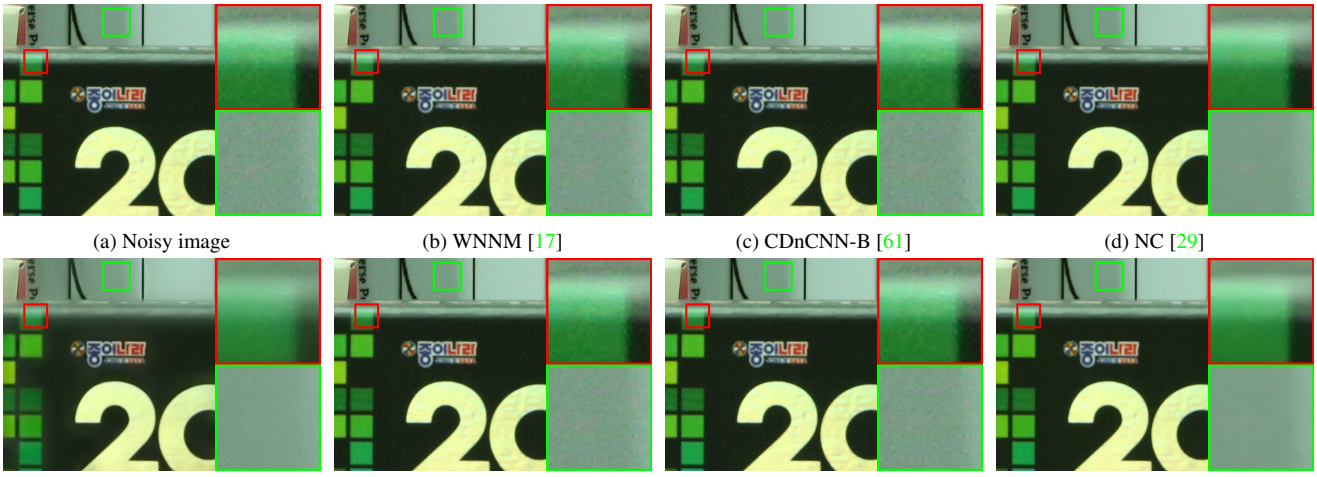
Figure 3: Denoising results of another NC12 image by different methods.



(a) Noisy image (b) BM3D [12] (c) CDnCNN-B [61] (d) NC [29]

(e) NI [2] (f) MCWNNM [59] (g) TWSC [58] (h) CBDNet

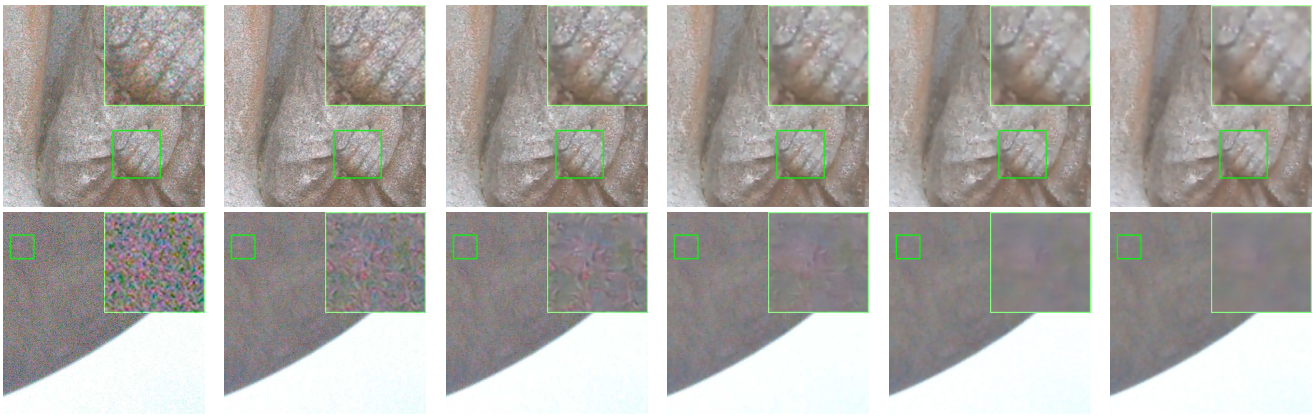
Figure 4: Denoising results of a DND image by different methods.



(a) Noisy image (b) WNNM [17] (c) CDnCNN-B [61] (d) NC [29]

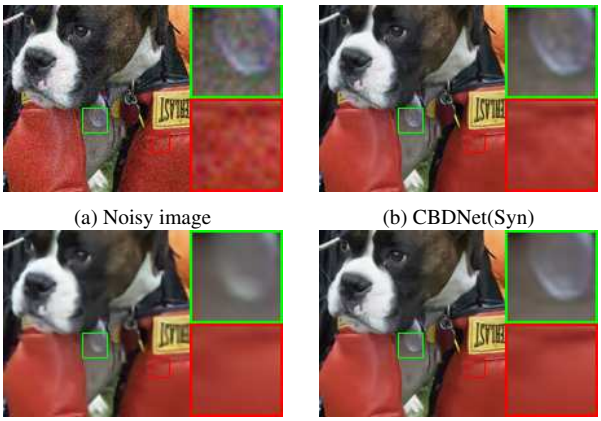
(e) NI [2] (f) MCWNNM [59] (g) TWSC [58] (h) CBDNet

Figure 5: Denoising results of a Nam image by different methods.



(a) Noisy (b) $\gamma = 0.4$ (c) $\gamma = 0.7$ (d) $\gamma = 1.0$ (e) $\gamma = 1.3$ (f) $\gamma = 1.6$

Figure 6: Results by interactive image denoising on two DND images.



(a) Noisy image (b) CBDNet(Syn) (c) CBDNet(Real) (d) CBDNet(All)

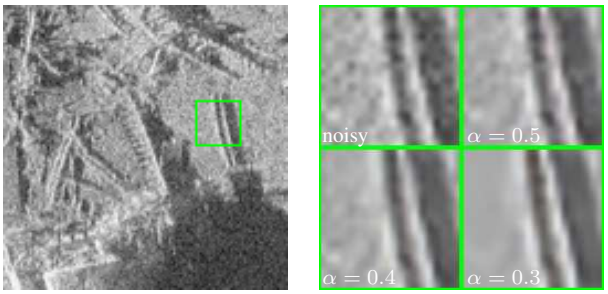
Figure 7: Denoising results of CBDNet trained by different data.

noise-free images. In comparison, CBDNet(All) is effective in removing real noise while preserving sharp edges. Also quantitative results of the three models on DND are shown in Table 1. CBDNet(All) obtains better PSNR/SSIM results than CBDNet(Syn) and CBDNet(Real).

Asymmetric loss. Fig. 8 compares the denoising results of CBDNet with different α values, *i.e.*, $\alpha = 0.5, 0.4$ and 0.3. CBDNet imposes equal penalty to under-estimation and over-estimation errors when $\alpha = 0.5$, and more penalty is imposed on under-estimation error when $\alpha < 0.5$. It can be seen that smaller α (*i.e.*, 0.3) is helpful in improving the generalization ability of CBDNet to unknown real noise.

4.5. Interactive Image Denoising

Given the estimated noise level map $\hat{\sigma}(\mathbf{y})$, we introduce a coefficient $\gamma\,(> 0)$ to interactively modify $\hat{\sigma}(\mathbf{y})$ to $\hat{\varrho} = \gamma \cdot \hat{\sigma}(\mathbf{y})$. By allowing the user to adjust γ, the non-blind denoising subnetwork takes $\hat{\varrho}$ and the noisy image as input to obtain denoising result. Fig. 6 presents two real noisy DND images as well as the results obtained using different γ values. By specifying $\gamma = 0.7$ to the first image and $\gamma = 1.3$ to the second, CBDNet can achieve the results with better visual quality in preserving detailed textures and removing sophisticated noise, respectively. Such interactive scheme can thus provide a convenient means for adjusting the denosing results in practical scenario.



(a) Noisy image (b) Denoised patches

Figure 8: Denoising results of CBDNet with different α values

5. Conclusion

We presented a CBDNet for blind denoising of real-world noisy photographs. The main findings of this work are two-fold. First, realistic noise model, including heterogenous Gaussian and ISP pipeline, is critical in making the learned model from synthetic images be applicable to real-world noisy photographs. Second, the denoising performance of a network can be boosted by incorporating both synthetic and real noisy images in training. Moreover, by introducing a noise estimation subnetwork into CBDNet, we were able to utilize asymmetric loss to improve its generalization ability to real-world noise, and perform interactive denoising conveniently.

6. Acknowledgements

This work is supported by NSFC (grant no. 61671182, 61872118, 61672446) and HK RGC General Research Fund (PolyU 152216/18E).

References

[1] Abdelrahman Abdelhamed, Stephen Lin, and Michael S Brown. A high-quality denoising dataset for smartphone cameras. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1692–1700, 2018. 2, 5

[2] Neatlab ABSoft. Neat image. https://ni.neatvideo.com/home. 6, 7

[3] Michal Aharon, Michael Elad, and Alfred Bruckstein. K-svd: An algorithm for designing overcomplete dictionaries for sparse representation. 2006. 1, 6

[4] Josue Anaya and Adrian Barbu. Renoir - a dataset for real low-light noise image reduction. 2014. 5

[5] Saeed Anwar, Cong Phuoc Huynh, and Fatih Murat Porikli. Chaining identity mapping modules for image denoising. *CoRR*, abs/1712.02933, 2017. 6

[6] Harold Christopher Burger, Christian J. Schuler, and Stefan Harmeling. Image denoising: Can plain neural networks compete with bm3d? *2012 IEEE Conference on Computer Vision and Pattern Recognition*, pages 2392–2399, 2012. 2, 6

[7] Vladimir Bychkovsky, Sylvain Paris, Eric Chan, and Frédo Durand. Learning photographic global tonal adjustment with a database of input / output image pairs. In *The Twenty-Fourth IEEE Conference on Computer Vision and Pattern Recognition*, 2011. 5

[8] Priyam Chatterjee and Peyman Milanfar. Is denoising dead? *IEEE Transactions on Image Processing*, 19:895–911, 2010. 1

[9] Guangyong Chen, Fengyuan Zhu, and Pheng-Ann Heng. An efficient statistical method for image noise level estimation. *2015 IEEE International Conference on Computer Vision (ICCV)*, pages 477–485, 2015. 3, 6

[10] Jingwen Chen, Jiawei Chen, Hongyang Chao, and Ming Yang. Image blind denoising with generative adversarial network based noise modeling. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3155–3164, 2018. 6

[11] Yunjin Chen and Thomas Pock. Trainable nonlinear reaction diffusion: A flexible framework for fast and effective image restoration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39:1256–1272, 2017. 1, 2, 6

[12] Kostadin Dabov, Alessandro Foi, Vladimir Katkovnik, and Karen O. Egiazarian. Color image denoising via sparse 3d collaborative filtering with grouping constraint in luminance-chrominance space. *2007 IEEE International Conference on Image Processing*, 1:I – 313–I – 316, 2007. 1, 2, 6, 7

[13] Weisheng Dong, Lei Zhang, Guangming Shi, and Xin Li. Nonlocally centralized sparse representation for image restoration. *IEEE Transactions on Image Processing*, 22:1620–1630, 2013. 6

[14] Alessandro Foi, Mejdi Trimeche, Vladimir Katkovnik, and Karen O. Egiazarian. Practical poissonian-gaussian noise modeling and fitting for single-image raw-data. *IEEE Transactions on Image Processing*, 17:1737–1754, 2008. 2, 3, 4

[15] Zheng Gong, Zuowei Shen, and Kim-Chuan Toh. Image restoration with mixed or unknown noises. *Multiscale Modeling and Simulation*, 12:458–487, 2014. 3

[16] Michael D. Grossberg and Shree K. Nayar. Modeling the space of camera response functions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26:1272–1282, 2004. 4, 5

[17] Shuhang Gu, Lei Zhang, Wangmeng Zuo, and Xiangchu Feng. Weighted nuclear norm minimization with application to image denoising. *2014 IEEE Conference on Computer Vision and Pattern Recognition*, pages 2862–2869, 2014. 1, 6, 7

[18] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. *2015 IEEE International Conference on Computer Vision (ICCV)*, pages 1026–1034, 2015. 5

[19] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 770–778, 2016. 2

[20] Youngbae Hwang, Jun-Sik Kim, and In-So Kweon. Difference-based image noise modeling using skellam distribution. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34:1329–1341, 2012. 3

[21] Sergey Ioffe and Christian Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *ICML*, 2015. 2

[22] Viren Jain and H. Sebastian Seung. Natural image denoising with convolutional networks. In *NIPS*, 2008. 2

[23] Justin Johnson, Alexandre Alahi, and Li Fei-Fei. Perceptual losses for real-time style transfer and super-resolution. In *European Conference on Computer Vision*, pages 694–711. Springer, 2016. 5

[24] Charles Kervrann, Jérôme Boulanger, and Pierrick Coupé. Bayesian non-local means filter, image redundancy and adaptive dictionaries for noise removal. In *SSVM*, 2007. 3

[25] Seon Joo Kim, Hai Ting Lin, Zheng Lu, Sabine Süsstrunk, Stephen Lin, and Michael S. Brown. A new in-camera imaging model for color computer vision and its application. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34:2289–2302, 2012. 3

[26] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *CoRR*, abs/1412.6980, 2014. 5

[27] Marc Lebrun, Antoni Buades, and Jean-Michel Morel. A nonlocal bayesian image denoising algorithm. *SIAM Journal on Imaging Sciences*, 6(3):1665–1688, 2013. 3

[28] Marc Lebrun, Miguel Colom, and Jean-Michel Morel. Multiscale image blind denoising. *IEEE Transactions on Image Processing*, 24:3149–3161, 2015. 3, 6

[29] Marc Lebrun, Miguel Colom, and Jean-Michel Morel. The noise clinic: a blind image denoising algorithm. *IPOL Journal*, 5:1–54, 2015. 2, 3, 5, 6, 7

[30] Jaakko Lehtinen, Jacob Munkberg, Jon Hasselgren, Samuli Laine, Tero Karras, Miika Aittala, and Timo Aila. Noise2noise: Learning image restoration without clean data. *arXiv preprint arXiv:1803.04189*, 2018. 2

[31] Anat Levin, Boaz Nadler, Frédo Durand, and William T. Freeman. Patch complexity, finite pixel correlations and optimal denoising. In *ECCV*, 2012. 1

[32] Ce Liu, Richard Szeliski, Sing Bing Kang, C. Lawrence Zitnick, and William T. Freeman. Automatic estimation and removal of noise from a single image. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30:299–314, 2008. 2, 3

[33] Pengju Liu, Hongzhi Zhang, Kai Zhang, Liang Lin, and Wangmeng Zuo. Multi-level wavelet-cnn for image restoration. *CoRR*, abs/1805.07071, 2018. 2

[34] Xinhao Liu, Masayuki Tanaka, and Masatoshi Okutomi. Single-image noise level estimation for blind denoising. *IEEE Transactions on Image Processing*, 22:5226–5237, 2013. 3

[35] Xinhao Liu, Masayuki Tanaka, and Masatoshi Okutomi. Practical signal-dependent noise parameter estimation from a single noisy image. *IEEE Transactions on Image Processing*, 23:4361–4371, 2014. 3

[36] Kede Ma, Zhengfang Duanmu, Qingbo Wu, Zhou Wang, Hongwei Yong, Hongliang Li, and Lei Zhang. Waterloo exploration database: New challenges for image quality assessment models. *IEEE Transactions on Image Processing*, 26:1004–1016, 2016. 5

[37] Henrique S Malvar, Li-wei He, and Ross Cutler. High-quality linear interpolation for demosaicing of bayer-patterned color images. In *Acoustics, Speech, and Signal Processing, 2004. Proceedings.(ICASSP'04). IEEE International Conference on*, volume 3, pages iii–485. IEEE, 2004. 4

[38] Xiao-Jiao Mao, Chunhua Shen, and Yu-Bin Yang. Image restoration using very deep convolutional encoder-decoder networks with symmetric skip connections. In *NIPS*, 2016. 1, 2

[39] Charles H. Martin and Michael W. Mahoney. Rethinking generalization requires revisiting old ideas: statistical mechanics approaches and complex learning behavior. *CoRR*, abs/1710.09553, 2017. 1, 2, 3

[40] D. Martin, C. Fowlkes, D. Tal, and J. Malik. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In *Proc. 8th Int'l Conf. Computer Vision*, volume 2, pages 416–423, July 2001. 5

[41] Ben Mildenhall, Jonathan T Barron, Jiawen Chen, Dillon Sharlet, Ren Ng, and Robert Carroll. Burst denoising with kernel prediction networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2502–2510, 2018. 4

[42] Vinod Nair and Geoffrey E. Hinton. Rectified linear units improve restricted boltzmann machines. In *ICML*, 2010. 4

[43] 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. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1683–1691, 2016. 2, 3, 5, 6

[44] Alberto Ortiz and Gabriel Oliver. Radiometric calibration of ccd sensors: dark current and fixed pattern noise estimation. *IEEE International Conference on Robotics and Automation, 2004. Proceedings. ICRA '04. 2004*, 5:4730–4735 Vol.5, 2004. 2

[45] Tobias Plotz and Stefan Roth. Benchmarking denoising algorithms with real photographs. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2750–2759, 2017. 1, 2, 5, 6

[46] Javier Portilla. Blind non-white noise removal in images using gaussian scale mixtures in the wavelet domain. In *Benelux Signal Processing Symposium*, 2004. 3

[47] Javier Portilla. Full blind denoising through noise covariance estimation using gaussian scale mixtures in the wavelet domain. *2004 International Conference on Image Processing, 2004. ICIP '04.*, 2:1217–1220 Vol.2, 2004. 3

[48] Stanislav Pyatykh, Jürgen Hesser, and Lei Zheng. Image noise level estimation by principal component analysis. *IEEE Transactions on Image Processing*, 22:687–699, 2013. 3

[49] Tamer F. Rabie. Robust estimation approach for blind denoising. *IEEE Transactions on Image Processing*, 14:1755–1765, 2005. 2, 3

[50] Yaniv Romano, Michael Elad, and Peyman Milanfar. The little engine that could: Regularization by denoising (red). *CoRR*, abs/1611.02862, 2016. 1

[51] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 234–241. Springer, 2015. 4

[52] Stefan Roth and Michael J. Black. Fields of experts: a framework for learning image priors. *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, 2:860–867 vol. 2, 2005. 6

[53] Uwe Schmidt and Stefan Roth. Shrinkage fields for effective image restoration. *2014 IEEE Conference on Computer Vision and Pattern Recognition*, pages 2774–2781, 2014. 1, 2

[54] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014. 5

[55] Ying Tai, Jian Yang, Xiaoming Liu, and Chunyan Xu. Memnet: A persistent memory network for image restoration. *2017 IEEE International Conference on Computer Vision (ICCV)*, pages 4549–4557, 2017. 2

[56] Andrea Vedaldi and Karel Lenc. Matconvnet - convolutional neural networks for matlab. In *ACM Multimedia*, 2015. 5

[57] Junyuan Xie, Linli Xu, and Enhong Chen. Image denoising and inpainting with deep neural networks. In *NIPS*, 2012. 2

[58] Jun Xu, Lei Zhang, and David Zhang. A trilateral weighted sparse coding scheme for real-world image denoising. In *European Conference on Computer Vision*, 2018. 3, 6, 7

[59] Jun Xu, Lei Zhang, David Zhang, and Xiangchu Feng. Multi-channel weighted nuclear norm minimization for real color image denoising. *2017 IEEE International Conference on Computer Vision (ICCV)*, pages 1105–1113, 2017. 3, 6, 7

[60] Dong Yang and Jian Sun. Bm3d-net: A convolutional neural network for transform-domain collaborative filtering. *IEEE Signal Processing Letters*, 25:55–59, 2018. 2

[61] Kai Zhang, Wangmeng Zuo, Yunjin Chen, Deyu Meng, and Lei Zhang. Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising. *IEEE Transactions on Image Processing*, 26:3142–3155, 2017. 1, 2, 3, 4, 6, 7

[62] Kai Zhang, Wangmeng Zuo, and Lei Zhang. Ffdnet: Toward a fast and flexible solution for cnn based image denoising. *CoRR*, abs/1710.04026, 2017. 1, 2, 4, 6, 7

[63] Fengyuan Zhu, Guangyong Chen, and Pheng-Ann Heng. From noise modeling to blind image denoising. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 420–429, 2016. 3

[64] Daniel Zoran and Yair Weiss. From learning models of natural image patches to whole image restoration. *2011 International Conference on Computer Vision*, pages 479–486, 2011. 6