When AWGN-Based Denoiser Meets Real Noises

Yuqian Zhou,[1] Jianbo Jiao,[2*] Haibin Huang,[3] Yang Wang,[4]
Jue Wang,[3] Honghui Shi,[1] Thomas Huang[1]
[1]IFP Group, UIUC, [2]University of Oxford, [3]Megvii Research, [4]Stony Brook University
{yuqian2, t-huang1}@illinois.edu, {jiaojianbo.i, shihonghui3}@gmail.com,
{huanghaibin, wangjue}@megvii.com, wang33@cs.stonybrook.edu

Abstract

Discriminative learning based image denoisers have achieved promising performance on synthetic noises such as Additive White Gaussian Noise (AWGN). The synthetic noises adopted in most previous work are pixel-independent, but real noises are mostly spatially/channel-correlated and spatially/channel-variant. This domain gap yields unsatisfied performance on images with real noises if the model is only trained with AWGN. In this paper, we propose a novel approach to boost the performance of a real image denoiser which is trained only with synthetic pixel-independent noise data dominated by AWGN. First, we train a deep model that consists of a noise estimator and a denoiser with mixed AWGN and Random Value Impulse Noise (RVIN). We then investigate Pixel-shuffle Down-sampling (PD) strategy to adapt the trained model to real noises. Extensive experiments demonstrate the effectiveness and generalization of the proposed approach. Notably, our method achieves state-of-the-art performance on real sRGB images in the DND benchmark among models trained with synthetic noises. Codes are available at https://github.com/yzhouas/PD-Denoising-pytorch.

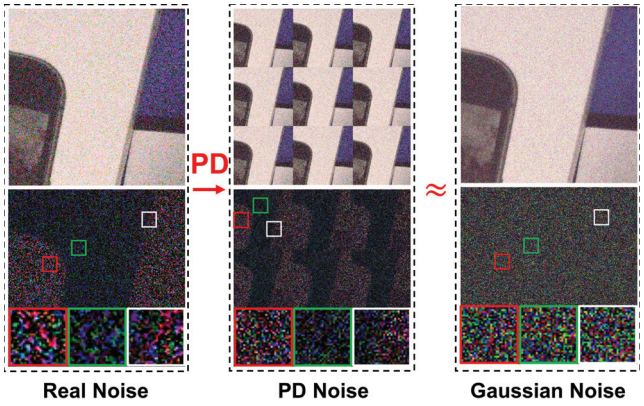


Figure 1: Basic idea of the proposed adaptation method: Pixel-shuffle Down-sampling (PD). Spatially-correlated real noise (Left) is broken into spatially-variant pixel-independent noise (Middle) to approximate spatially-variant Gaussian noise (Right). Then an AWGN-based denoiser can be applied to such real noise accordingly.

Introduction

As a fundamental task in image processing and computer vision, image denoising has been extensively explored in the past several decades even for downstream applications (Zhou, Liu, and Huang 2018; Wang et al. 2019). Traditional methods including the ones based on image filtering (Dabov et al. 2008), low rank approximation (Gu et al. 2014; Xu et al. 2017; Yair and Michaeli 2018), sparse coding (Elad and Aharon 2006), and image prior (Ulyanov, Vedaldi, and Lempitsky 2017) have achieved satisfactory results on synthetic noise such as Additive White Gaussian Noise (AWGN). Recently, deep CNN has been applied to this task, and discriminative-learning-based methods such as DnCNN (Zhang et al. 2017a) outperform most traditional methods on AWGN denoising.

Unfortunately, while these learning-based methods work well on the same type of synthetic noise that they were trained on, their performance degrades rapidly on real images, showing poor generalization ability in real world applications. This indicates that these data-driven denoising models are highly domain-specific and non-flexible to transfer to other noise types beyond AWGN. To improve model flexibility, the recently-proposed FFDNet (Zhang, Zuo, and Zhang 2018) trains a conditional non-blind denoiser with a manually adjusted noise-level map. By giving high-valued uniform maps to FFDNet, only over-smoothed results can be obtained in real image denoising. Therefore, blind denoising of real images is still very challenging due to the lack of accurate modeling of real noise distribution. These unknown real-world noises are much more complex than pixel-independent AWGN. They can be spatially-variant, spatially-correlated, signal-dependent, and even device-dependent.

To better address the problem of real image denoising, current attempts can be roughly divided into the following categories: (1) realistic noise modeling (Shi Guo 2018; Brooks et al. 2019; Abdelhamed, Timofte, and Brown 2019),

*Corresponding author

(2) noise profiling such as multi-scale (Lebrun, Colom, and Morel 2015a; Yair and Michaeli 2018), multi-channel (Xu et al. 2017) and regional based (Liu et al. 2017) settings, and (3) data augmentation techniques such as the adversarial-learning-based ones (Chen et al. 2018). Among them, CBDNet (Shi Guo 2018) achieves good performance by modeling the realistic noise using the in-camera pipeline model proposed in (Liu et al. 2008). It also trains an explicit noise estimator and sets a larger penalty for under-estimated noise. The network is trained on both synthetic and real noises, but it still cannot fully characterize real noises. Brooks et al. (Brooks et al. 2019) used prior statistics stored in the raw data of DND to augment the synthetic RGB data, but it does not prove the generalization of the model on other real noises.

In this work, from a novel viewpoint of real image blind denoising, we seek to adapt a learning-based denoiser trained on pixel-independent synthetic noises to unknown real noises. As shown in Figure 1, we assume that real noises differ from pixel-independent synthetic noises dominantly in *spatial/channel-variance and correlation* (Stanford 2015). This difference results from in-camera pipeline like demosaicing (Zhou et al. 2019). Based on this assumption, we first propose to train a basis denoising network using mixed AWGN and RVIN. Our flexible basis net consists of an explicit noise estimator followed by a conditional denoiser. We demonstrate that this fully-convolutional nets are actually efficient in coping with pixel-independent spatially/channel-variant noises. Second, we propose a simple yet effective adaptation strategy, Pixel-shuffle Down-sampling(PD), which employs the divide-and-conquer idea to handle real noises by breaking down the spatial correlation.

In summary, our main contributions include:

- We propose a new flexible deep denoising model (trained with AWGN and RVIN) for both blind and non-blind image denoising. We also demonstrate that such fully convolutional models trained on spatially-invariant noises can handle *spatially-variant noises*.
- We adapt the AWGN-RVIN-trained deep denoiser to real noises by applying a novel strategy called Pixel-shuffle Down-sampling (PD). *Spatially-correlated noises* are broken down to *pixel-wise independent noises*. We examine and overcome the proposed domain gap to boost real denoising performance.
- The proposed method achieves state-of-the-art performance on DND benchmark and other real noisy RGB images among models trained only with synthetic noises. Note that our model does not use any images or prior meta-data from real noise datasets. We also show that with the proposed PD strategy, the performance of some other existing denoising models can also be boosted.

Related Work

Discriminative Learning based Denoiser. Denoising methods based on CNNs have achieved impressive performance on removing synthetic Gaussian noise. Burger et al. (Burger, Schuler, and Harmeling 2012) proposed to apply multi-layer perceptron (MLP) to denoising task. In (Chen and Pock 2017), Chen et al. proposed a trainable nonlinear reaction diffusion (TNRD) model for Gaussian noise removal at different level. DnCNN (Zhang et al. 2017a) was the first to propose a blind Gaussian denoising network using deep CNNs. It demonstrated the effectiveness of residual learning and batch normalization. More network structures like dilated convolution (Zhang et al. 2017b), autoencoder with skip connection (Mao, Shen, and Yang 2016), ResNet (Ren, El-Khamy, and Lee 2018), recursively branched deconvolutional network (RBDN) (Santhanam, Morariu, and Davis 2017) were proposed to either enlarge the receptive field or balance the efficiency. Recently some interests are put into combining image denoising with high-level vision tasks like classification and segmentation. Liu et al. (Liu et al. 2017) applied segmentation to enhance the denoising performance on different regions. Similar class-aware work were developed in (Niknejad, Bioucas-Dias, and Figueiredo 2017). Due to domain-specific training and deficient realistic noise data, those deep models are not robust enough on realistic noises. In recently proposed FFDNet (Zhang, Zuo, and Zhang 2018), the author proposed a non-blind denoising by concatenating the noise level as a map to the noisy image. By manually adjusting noise level to a higher value, FFDNet demonstrates a spatial-invariant denoising on realistic noises with over-smoothed details.

Blind Denoising on Real Noisy Images. Real noises of CCD cameras are complicated and are related to optical sensors and in-camera process. Specifically, multiple noise sources like photon noise, read-out noise etc. and processing including demosaicing, color and gamma transformation introduce the main characteristics of real noises: spatial/channel correlation, variance, and signal-dependence. To approximate real noise, multiple types of synthetic noise are explored in previous work, including Gaussian-Poisson (Foi et al. 2008; Liu, Tanaka, and Okutomi 2014), Gaussian Mixture Model (GMM) (Zhu, Chen, and Heng 2016), in-camera process simulation (Liu et al. 2008; Shi Guo 2018) and GAN-generated noises (Chen et al. 2018), to name a few. CBDNet (Shi Guo 2018) first simulated real noise and trained a subnetwork for noise estimation, in which spatial-variance noise is represented as spatial maps. Besides, multi-channel (Xu et al. 2017; Shi Guo 2018) and multi-scale (Lebrun, Colom, and Morel 2015a; Yu and Koltun 2015) strategy were also investigated for adaptation. Different from all the aforementioned works which focus on directly synthesizing or simulating noises for training, in this work, we apply AWGN-RVIN model and focus on pixel-shuffle adaptation strategy to fill in the gap between pixel-independent synthetic and pixel-correlated real noises.

Methodology

Basis Noise Model

The basis noise model is mixed AWGN-RVIN. Noises in sRGB images are no longer approximated Gaussian-Poisson Noises as in the raw sensor data mainly due to gamma transform, demosaicing, and other interpolations etc.. In Figure 2, we follow (Liu et al. 2008) pipeline to synthesize noisy im-

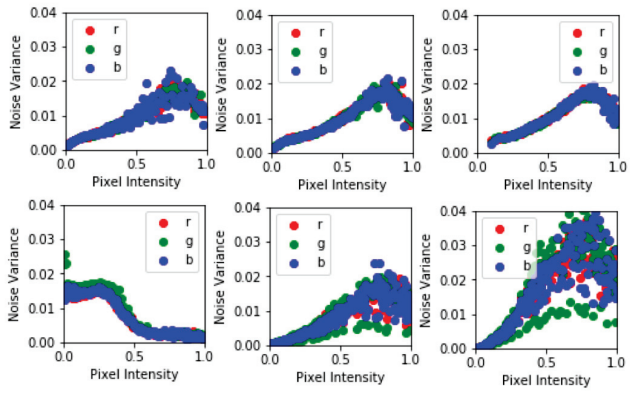


Figure 2: Noise Level Function (NLFs) (noise variance as a function of image intensity) before (first row) and after (second row) gamma transform and demosaicing. Gamma factor is 0.39, 1.38 and 2.31 from the left to right column.

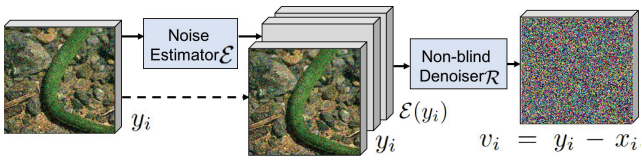


Figure 3: Structure of the proposed blind denoising model. It consists of a noise estimator $\mathcal{E}$ and a follow-up non-blind denoiser $\mathcal{R}$. The model aims to jointly learn the image residual.

ages, and plot the Noise Level Functions (NLFs) (noise variance as a function of image intensity) before (first row) and after (second row) the Gamma Correction transform and demosaicing. From left to right, the Gamma factor increases. It shows that in RGB images, clipping effect and other nonlinear transforms will greatly influence the originally linear noise variance-intensity relationship in raw sensor data, even change the noise mean. Tough complicated, for a more general case than Gaussian-Poisson noises of modeling different nonlinear transforms, real noises in RGB can still be locally approximated as AWGN (Zhang, Zuo, and Zhang 2018; Lee 1980; Xu, Zhang, and Zhang 2018). In this paper, we thus assume the RGB noises to be approximated spatially-variant and spatially-correlated AWGN.

Adding RVIN for training aims at explicitly resolving the defective pixels caused by dead pixels of camera hardware or long exposure frequently appearing in most night-shot images. We generate AWGN, RVIN and mixed AWGN-RVIN following PGB(Xu et al. 2016).

Basis Model Structure

The architecture of the proposed basis model is illustrated in Figure 3. The proposed blind denoising model $\mathcal{G}$ consists of a noise estimator $\mathcal{E}$ and a follow-up non-blind denoiser $\mathcal{R}$. Given a noisy observation $y_i = \mathcal{F}(x_i)$, where $\mathcal{F}$ is the noise synthetic process, and x_i is the noise-free image, the model aims to jointly learn the residual $\mathcal{G}(y_i) \approx v_i = y_i - x_i$, and it is trained on paired synthetic data (y_i, v_i). Specifically, the noise estimator outputs $\mathcal{E}(y_i)$ consisting of six pixel-wise noise-level maps that correspond to two noise types, i.e., AWGN and RVIN, across three channels (R, G, B). Then y_i is concatenated with the estimated noise level maps $\mathcal{E}(y_i)$ and fed into the non-blind denoiser $\mathcal{R}$. The denoiser then outputs the noise residual $\mathcal{G}(y_i) = \mathcal{R}(y_i, \mathcal{E}(y_i))$. Three objectives are proposed to supervise the network training, including the noise estimation ($\mathcal{L}_e$), blind ($\mathcal{L}_b$) and non-blind ($\mathcal{L}_{nb}$) image denoising objectives, defined as,

$$\mathcal{L}_e = \frac{1}{2N}\sum_{i=1}^{N} ||\mathcal{E}(y_i;\Theta_E) - e_i||_F^2, \tag{1}$$

$$\mathcal{L}_b = \frac{1}{2N}\sum_{i=1}^{N} ||\mathcal{R}(y_i, \mathcal{E}(y_i;\Theta_E);\Theta_R) - v_i||_F^2, \tag{2}$$

$$\mathcal{L}_{nb} = \frac{1}{2N}\sum_{i=1}^{N} ||\mathcal{R}(y_i, e_i;\Theta_R) - v_i||_F^2, \tag{3}$$

where Θ_E and Θ_R are the trainable parameters of $\mathcal{E}$ and $\mathcal{R}$. e_i is the ground truth noise level maps for y_i, consisting of e_{iAWGN} and e_{iRVIN}. For AWGN, e_{iAWGN} is represented as the even maps filled with the same standard deviation values ranging from 0 to 75 across R,G,B channels. For RVIN, e_{iRVIN} is represented as the maps valued with the corrupted pixels ratio with upper-bound set to 0.3. We further normalize e_i to range [0,1]. Then the full objective can be represented as a weighted sum of the above three losses,

$$\mathcal{L} = \alpha\mathcal{L}_e + \beta\mathcal{L}_b + \gamma\mathcal{L}_{nb}, \tag{4}$$

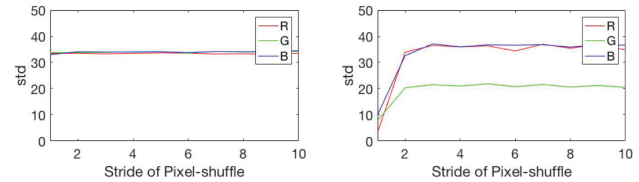
in which α, β and γ are hyper-parameters to balance the losses, and we set them to be equal for simplicity.

The proposed model structure can perform both blind and non-blind denoising simultaneously, and the model is more flexible in interactive denoising and result adjustment. Explicit noise estimation also benefits noise modeling and disentanglement.

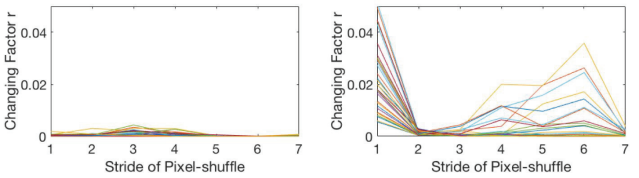
Pixel-shuffle Down-sampling (PD) Adaptation

Pixel-shuffle Down-sampling. Pixel-shuffle (Shi et al. 2016) down-sampling is defined to create the mosaic by sampling the images with stride s. Compared to other down-sampling methods like linear interpolation, bi-cubic interpolation, and pixel area relation, the pixel-shuffle and nearest-neighbour down-sampling on noisy image would not influence the real noise distribution. Besides, pixel-shuffle also benefits image recovery by preserving the original pixels from the images compared to others. These two advantages yield the two stages of PD strategy: adaptation and refinement.

Adaptation. Learning-based denoiser trained on AWGN is not robust enough to real noises due to domain difference. To adapt the noise model to real noise, here we briefly analyze and justify our assumption on the difference between real noises and Gaussian noise: spatial/channel variance and correlation.



(a) As the stride increases, Left: Estimated noise level on AWGN-corrupted image. Right: Estimated noise level on real noisy images.



(b) Left: Changing factor r_s on AWGN-corrupted images of CBSD68 and Right: on real noisy images of DND. Different color lines represent different image samples.

Figure 4: Influence of Pixel-shuffle on noise patterns and noise estimation algorithms.

Suppose a noise estimator is robust, which means it can accurately estimate the exact noise level, for a single AWGN-corrupted image, pixel-shuffle down-sampling will neither influence the AWGN variance nor the estimation values, when the sample stride is small enough to preserve the textural structures. When extending it to real noise case, we have an interesting hypothesis: as we increase the sample stride of pixel-shuffle, the estimation values of specific noise estimators will first fluctuate and then keep steady for a couple of stride increment. This assumption is feasible because pixel-shuffle will break down the spatial-correlated noise patterns to pixel-independent ones, which can be approximated as spatial-variant AWGN and adapted to those estimators.

We justify this hypothesis on both (Liu, Tanaka, and Okutomi 2013) and our proposed pixel-wise estimator. As shown in Figure 1, we randomly cropped a patch of size 200×200 from a random noisy image y in SIDD(Abdelhamed, Lin, and Brown 2018). We add AWGN with $std = 35$ to its noise-free ground truth x. After pixel-shuffling both y and AWGN-corrupted x, starting from stride $s = 2$, the noise pattern of y demonstrates expected pixel independence. Using (Liu, Tanaka, and Okutomi 2013), the estimation result for x is unchanged in Figure 4 (a) (Left), but the one for y in Figure 4 (a) (Right) first increases and begins to keep steady after stride $s = 2$. It is consistent with the visual pattern and our hypothesis.

One assumption of (Liu, Tanaka, and Okutomi 2013) is that the noise is additive and evenly distributed across the image. For spatial-variant signal-dependent real noises, our pixel-wise estimator has its superiority. To make statistics of spatial-variant noise estimation values, we extract the three AWGN channels of noise map $\mathcal{E}_{AWGN}(y_i) \in R^{W \times H \times 3}$, where W and H are width and height of the input image, and compute the normalized 10-bin histograms $h_s \in R^{10 \times 3}$ across each channel when the stride is s. We introduce the changing factor r_s to monitor the noise map distribution changes as the stride s increases,

$$r_s = E_c ||h_{sc} - h_{(s+1)c}||_2^2, \tag{5}$$

where c is the channel index. We then investigate the difference of r_s sequence between AWGN and realistic noises. Specifically, we randomly select 50 images from CBSD68 (Roth and Black 2009) and add random-level AWGN to them. For comparison, we randomly pick up 50 image patches of 512×512 from DND benchmark. In Figure 4 (b), r_s sequence remains closed to zero for all AWGN-currupted images (Left figure), while for real noises r_α demonstrates an abrupt drop when $s = 2$. It indicates that the spatial-correlation has been broken from $s = 2$.

The above analysis inspires the proposed adaptation strategy based on pixel-shuffle. Intuitively, we aim at finding the smallest stride s to make the down-sampled spatial-correlated noises match the pixel-independent AWGN. Thus we keep increasing the stride s until r_s drops under a threshold τ. We run the above experiments on CBSD68 for 100 iterations to select the proper generalized threshold τ. After averaging the maximum r of each iteration, we empirically set $\tau = 0.008$.

PD Refinement. Figure 5 shows the proposed Pixel-shuffle Down-sampling (PD) refinement strategy: (1) Compute the smallest stride s, which is 2 in this example and more digital camera image cases, to match AWGN following the adaptation process, and pixel-shuffle the image into mosaic y_s; (2) Denoise y_s using $\mathcal{G}$; (3) Refill each sub-image with noisy blocks separately and pixel-shuffle upsample them; (4) Denoise each refilled image again using $\mathcal{G}$ and average them to obtain the 'texture details' T; (5) Combine the over-smoothed 'flat regions' F to refine the final result.

As summarized in (Liu et al. 2008), the goals of noise removal include preserving texture details and boundaries, smoothing flat regions, and avoiding generating artifacts. Therefore, in the above step-(5), we propose to further refine the denoised image with the combination of 'texture details' T and 'flat regions' F. 'Flat regions' can be obtained from over-smoothed denoising results generated by lifting the noise estimation levels. In this work, given a noisy observation y, the refined noise maps are defined as,

$$\mathcal{E}(P\hat{D}(y))(i,j) = \max_{i,j} \mathcal{E}(PD(y))(i,j), i \in [1,W], j \in [1,H]. \tag{6}$$

Consequently, the 'flat region' is defined as $F = PU(\mathcal{R}(PD(y), \mathcal{E}(P\hat{D}(y))))$, where PD and PU are pixel-shuffle downsampling and upsampling. The final result is obtained by $kF + (1-k)T$.

Experiments

Implementation Details

In this work, the structures of the sub-network $\mathcal{E}$ and $\mathcal{R}$ follow DnCNN (Zhang et al. 2017a) of 5 layers and 20 layers. For grayscale image experiments, we also follow DnCNN to crop 50×50 patches from 400 images of size 180×180. For color image model, we crop 50×50 patches with stride

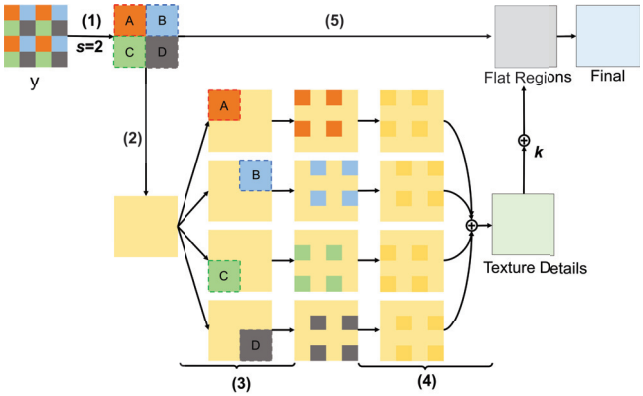


Figure 5: Pixel-shuffle Down-sampling (PD) refinement strategy with $s = 2$.

10 from 432 color images in the Berkeley segmentation dataset (BSD) (Roth and Black 2009). The training data ratio of single-type noises (either AWGN or RVIN) and mixed noises (AWGN and RVIN) is 1:1. During training, Adam optimizer is utilized and the learning rate is set to 10^{-3}, and batch size is 128. After 30 epochs, the learning rate drops to 10^{-4} and the training stops at epoch 50.

To evaluate the algorithm on synthetic noise (AWGN, mixed AWGN-RVIN and spatially-variant Gaussian), we utilize the benchmark data from BSD68, Set20 (Xu et al. 2016) and CBSD68 (Roth and Black 2009). For realistic noise, we test it on RNI15 (Online 2015a), DND benchmark (Plötz and Roth 2017), and self-captured night photos. We evaluate the performance of the algorithm in terms of PSNR and SSIM. Qualitative performance for denoising is also presented, with comparison to other state-of-the-arts.

Evaluation with Synthetic Noise

Table 1: Comparison of PSNR results on mixture of Gaussian noise (AWGN) and Impulse noise (RVIN) removal performance on Set20.

(σ, r)	BM3D	WNNM	PGB	DnCNN-B	Ours-NB	Ours-B
(10, 0.15)	25.18	25.41	27.17	32.09	**32.43**	32.37
(10, 0.30)	21.80	21.40	22.17	29.97	**30.47**	30.32
(20, 0.15)	25.13	23.57	26.12	29.52	**29.82**	29.76
(20, 0.30)	21.73	21.40	21.89	27.90	**28.41**	28.16

Mixed AWGN and RVIN. Our model follows similar structure of DnCNN and FFDNet (Zhang, Zuo, and Zhang 2018), so its performance on single-type AWGN removal is also similar to them. We thus evaluate our model on eliminating mixed AWGN and RVIN on Set20 as in (Xu et al. 2016). We also compare our method with other baselines, including BM3D (Dabov et al. 2006) and WNNM (Gu et al. 2014) which are non-blind Gaussian denoisers anchored with a specific noise level estimated by the approach provided in (Liu, Tanaka, and Okutomi 2013). Besides, we include the PGB (Xu et al. 2016) denoiser that is designed for mixed AWGN and RVIN. The result of the blind version of DnCNN-B, trained by the same strategy as our model, is also presented for reference. The comparison results are shown in Table 1, from which we can see the proposed method achieves the best performance. Compared to DnCNN-B, for complicated mixed noises, our model explicitly disentangles different noises. It benefits the conditional denoiser to differentiate mixed noises from other types.

Table 2: Comparison of PSNR results on Signal-dependent Noises on CBSD68.

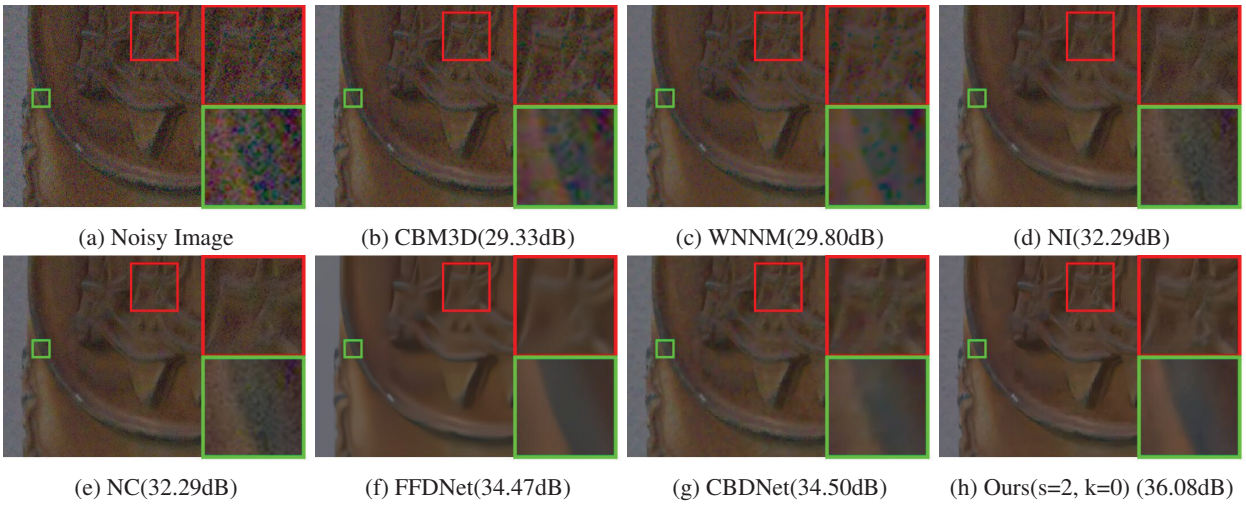
(σ_s, σ_c)	BM3D	FFDNet	DnCNN-B	CBDNet	Ours-B
(20, 10)	29.09	28.54	34.38	33.04	**34.75**
(20, 20)	29.08	28.70	**31.72**	29.77	31.32
(40, 10)	23.21	28.67	32.08	30.89	**32.12**
(40, 20)	23.21	28.80	30.32	28.76	**30.33**

Signal-dependent Spatially-variant Noise. We conduct experiments to examine the generalization ability of fully convolutional model on signal-dependent noise model (Shi Guo 2018; Foi et al. 2008; Liu, Tanaka, and Okutomi 2014). Given a clean image x, the noises in the noisy observation y contain both signal-dependent components with variance $x\sigma_s^2$ and independent components with variance σ_c^2. Table 2 shows that for non-blind model like BM3D and FFDNet, only scalar noise estimator (Liu, Tanaka, and Okutomi 2013) is applied, thus they cannot well cope with the spatially-variant cases. In this experiment, DnCNN-B is the original blind model trained on AWGN with σ ranged between 0 and 55. It shows that spatially-variant Gaussian noises can still be handled by fully convolutional model trained with spatially-invariant AWGN (Zhang, Zuo, and Zhang 2018). Compared to DnCNN-B, the proposed network explicitly estimates the pixel-wise map to make the model more flexible and possible for real noise adaptation.

Evaluation with Real RGB Noise

Qualitative Comparisons. Some qualitative denoising results on DND are shown in Figure 6. The compared results of DND are all directly obtained online from the original submissions of the authors. The methods we include for the comparison cover blind real denoisers (CBDNet, NI (Online 2015b) and NC (Lebrun, Colom, and Morel 2015b)), and non-blind Gaussian denoisers (CBM3D, WNNM (Gu et al. 2014), and FFDNet). From these example denoised results, we can observe that some of them are either noisy (as in WNNM), or spatially-invariantly over-smoothed (as in FFDNet). CBDNet performs better than others but it still suffers from blur edges and uncleaned background. Our proposed method (PD) achieves a better spatially-variant denoising performance by smoothing the background while preserving the textural details in a full blind setting.

Quantitative Results on DND Benchmark. The images in the DND benchmark are captured by digital camera and demosaiced from raw sensor data, so we simply set the stride number $s = 2$. We follow the submission guideline of DND dataset to evaluate our algorithm. Recently, many learning-based methods like Path-Restore (Yu et al. 2019),RID-



(a) Noisy Image (b) CBM3D(29.33dB) (c) WNNM(29.80dB) (d) NI(32.29dB)

(e) NC(32.29dB) (f) FFDNet(34.47dB) (g) CBDNet(34.50dB) (h) Ours(s=2, k=0) (36.08dB)

Figure 6: Denoising results on DND Benchmark. Red box indicates texture details while the green box background or edge.

Net (Anwar and Barnes 2019),WDnCNN (Zhao, Lam, and Lun 2019) and CBDNet, achieved promising performance on DND, but they are all finetuned on real noisy images, or use prior knowledge in the meta-data of DND (Brooks et al. 2019). For fair comparison, we select some representative conventional methods(MCWNNM, EPLL, TWSC, CBM3D), and learning-based methods trained only with synthetic noises. The results are shown in Table 3. Models trained on AWGN (DnCNN, TNRD, MLP) perform poorly on real RGB noises mainly due to the large gap between AWGN and real noise. CBDNet improves the results significantly by training the deep networks with artificial realistic noise model. Our AWGN-RVIN-trained model with PD refinement achieves much better results (+0.83dB) than CBDNet trained only with synthetic noises, and also boosts the performance of other AWGN-based methods (+PD). Compared to the base model, the proposed adaptation methods improve the performance on real noises by 5.8 dB. Note that our model is only trained on synthetic noises, and does not utilize any prior data of DND.

Ablation Study on Real RGB Noise

Adding RVIN. Training models with mixed AWGN and RVIN noises will benefit the removal of dead or over-exposure pixels in real images. For comparison, We train another model only with AWGN, and test it on real noisy night photos. An example utilizing the full pipeline is shown in Figure 7, in which it demonstrates the superiority of the existence of RVIN in the training data. Even though model trained with AWGN can also achieve promising denoising performance, it is not effective on dead pixels.

Stride Selection. We apply different stride numbers while refining the denoised results, and compare the visual quality in Figure 8 (a)(b). For arbitrary given sRGB images, the stride number can be computed using our adaptation algorithm with the assistance of noise estimator. In our experiments, the selected stride is the smallest s that $r_s < \tau$. Small stride number will treat large noise patterns as textures to preserve, as shown in Figure 8 (b). While using large stride number tends to break the textural structures and details. Interestingly, as shown in Figure 8 (b), the texture of the fabric is invisible while applying $s > 2$.

Table 3: Comparison of PSNR and SSIM on DND Benchmark. PD: Pixel-suffle Down-sampling Strategy. Among all models trained only with synthetic data.

Method	PSNR	SSIM
MCWNNM(Xu et al. 2017)	37.38	0.929
EPLL(Zoran and Weiss 2011)	33.51	0.824
TWSC(Xu, Zhang, and Zhang 2018)	37.93	0.940
MLP(Burger, Schuler, and Harmeling 2012)	34.23	0.833
TNRD(Chen and Pock 2017)	33.65	0.830
CBDNet(Syn)(Shi Guo 2018)	37.57	0.936
CBM3D(Dabov et al. 2008)	34.51	0.850
CBM3D(+PD)	*35.02*	*0.873*
CDnCNN-B(Zhang et al. 2017a)	32.43	0.790
CDnCNN-B(+PD)	*35.44*	*0.876*
FFDNet(Zhang, Zuo, and Zhang 2018)	34.40	0.847
FFDNet(+PD)	*37.56*	*0.931*
Our Base Model(No PD)	32.60	0.788
Ours(Full Pipeline)	**38.40**	**0.945**

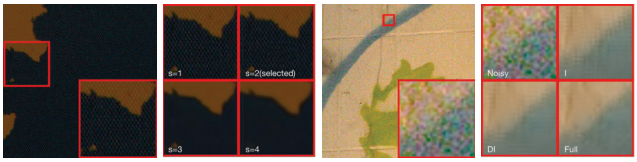
Table 4: Ablation study on refinement steps.

Model	(s=1)	(s=3, Full)	(s=2,I)	(s=2,DI)	(s=2,Full)
PSNR	32.60	37.90	37.00	37.20	**38.40**
SSIM	0.7882	0.9349	0.9339	0.9361	**0.9452**

Image Refinement Process. The ablation on the refinement steps is shown in Figure 8 (c)(d) and Table 4, in which we compare the denoised results of I (i.e. directly pixel-shuffling upsampling after step (2)), DI (i.e. denoising I using $\mathcal{G}$), and Full (i.e. the current whole pipeline). It shows that both I and DI will form additional visible artifacts, while the whole pipeline smooths out those artifacts and has the

(a) Noisy image (b) AWGN only (c) AWGN-RVIN

Figure 7: Denoised performance of models trained with AWGN in (b) and mixed AWGN-RVIN in (c). During testing, $k = 0$ and $s = 2$.



(a) Noisy image (b) Denoised. (c) Noisy Image (d) Denoised.

Figure 8: (a)(b):Denoised performance of different stride s when $k = 0$, and (c)(d): Ablation study on refinement. $s = 2$ and $k = 0$.

best visual quality.

Blending Factor k. Due to the ambiguity nature of fine texture and mid-frequent noises, human perception intervene on the denoising level is inevitable. k is this parameter introduced as a 'linear' adjustment of denoising level for a more flexible and interactive user operation. Using blending factor k is more stable and safe to preserve the spatially-variant details than directly adjusting the estimated noise level like CBDNet. In Figure 9, as k increases, the denoised results tend to be over-smoothed. This is suitable for images with more background patterns. However, smaller k will preserve more fine details which are applicable for images with more foreground objects. In most cases, users can simply set k to 0 to obtain the most detailed textures recovery and visually plausible results.

Conclusions

In this paper, we revisit the real image blind denoising from a new viewpoint. We assumed the realistic noises are spatially/channel -variant and correlated, and addressed adaptation from AWGN-RVIN noises to real noises. Specifically, we proposed an image blind and non-blind denoising network trained on AWGN-RVIN noise model. The network consists of an explicit multi-type multi-channel noise estimator and an adaptive conditional denoiser. To generalize the network to real noises, we investigated Pixel-shuffle Down-sampling (PD) refinement strategy. We showed qualitatively that PD behaves better in both spatially-variant denoising and details preservation. Results on DND benchmark and other realistic noisy images demonstrated the newly proposed model with the strategy are efficient in processing spatial/channel variance and correlation of real noises without explicit modeling.

(a) Noisy (b) 0 (c) 0.3 (d) 0.5 (e) 0.8 (f) 1

Figure 9: Ablation study on merging factor k, and $s = 2$.

References

Abdelhamed, A.; Lin, S.; and Brown, M. S. 2018. A high-quality denoising dataset for smartphone cameras. In *CVPR*.

Abdelhamed, A.; Timofte, R.; and Brown, M. S. 2019. Ntire 2019 challenge on real image denoising: Methods and results. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 0–0.

Anwar, S., and Barnes, N. 2019. Real image denoising with feature attention. *arXiv preprint arXiv:1904.07396*.

Brooks, T.; Mildenhall, B.; Xue, T.; Chen, J.; Sharlet, D.; and Barron, J. T. 2019. Unprocessing images for learned raw denoising. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 11036–11045.

Burger, H. C.; Schuler, C. J.; and Harmeling, S. 2012. Image denoising: Can plain neural networks compete with bm3d? In *CVPR*.

Chen, Y., and Pock, T. 2017. Trainable nonlinear reaction diffusion: A flexible framework for fast and effective image restoration. *IEEE transactions on pattern analysis and machine intelligence* 39(6):1256–1272.

Chen, J.; Chen, J.; Chao, H.; and Yang, M. 2018. Image blind denoising with generative adversarial network based noise modeling. In *CVPR*.

Dabov, K.; Foi, A.; Katkovnik, V.; and Egiazarian, K. 2006. Image denoising with block-matching and 3d filtering. In *Image Processing: Algorithms and Systems, Neural Networks, and Machine Learning*, volume 6064, 606414. International Society for Optics and Photonics.

Dabov, K.; Foi, A.; Katkovnik, V.; and Egiazarian, K. 2008. Image restoration by sparse 3d transform-domain collaborative filtering. In *Image Processing: Algorithms and Systems VI*, volume 6812, 681207. International Society for Optics and Photonics.

Elad, M., and Aharon, M. 2006. Image denoising via sparse and redundant representations over learned dictionaries. *IEEE Transactions on Image processing* 15(12):3736–3745.

Foi, A.; Trimeche, M.; Katkovnik, V.; and Egiazarian, K. 2008. Practical poissonian-gaussian noise modeling and fitting for single-image raw-data. *IEEE Transactions on Image Processing* 17(10):1737–1754.

Gu, S.; Zhang, L.; Zuo, W.; and Feng, X. 2014. Weighted nuclear norm minimization with application to image denoising. In *CVPR*.

Lebrun, M.; Colom, M.; and Morel, J.-M. 2015a. Multiscale image blind denoising. *IEEE Transactions on Image Processing* 24(10):3149–3161.

Lebrun, M.; Colom, M.; and Morel, J.-M. 2015b. The noise clinic: a blind image denoising algorithm. *Image Processing On Line* 5:1–54.

Lee, J.-S. 1980. Refined filtering of image noise using local statistics. Technical report, NAVAL RESEARCH LAB WASHINGTON DC.

Liu, C.; Szeliski, R.; Kang, S. B.; Zitnick, C. L.; and Freeman, W. T. 2008. Automatic estimation and removal of noise from a single image. *IEEE Trans. Pattern Anal. Mach. Intell.* 30(2):299–314.

Liu, D.; Wen, B.; Liu, X.; Wang, Z.; and Huang, T. S. 2017. When image denoising meets high-level vision tasks: A deep learning approach. *arXiv preprint arXiv:1706.04284*.

Liu, X.; Tanaka, M.; and Okutomi, M. 2013. Single-image noise level estimation for blind denoising. *IEEE transactions on image processing* 22(12):5226–5237.

Liu, X.; Tanaka, M.; and Okutomi, M. 2014. Practical signal-dependent noise parameter estimation from a single noisy image. *IEEE Transactions on Image Processing* 23(10):4361–4371.

Mao, X.; Shen, C.; and Yang, Y.-B. 2016. Image restoration using very deep convolutional encoder-decoder networks with symmetric skip connections. In *NeurIPS*.

Niknejad, M.; Bioucas-Dias, J. M.; and Figueiredo, M. A. 2017. Class-specific poisson denoising by patch-based importance sampling. *arXiv preprint arXiv:1706.02867*.

Online. 2015a. [online] available:. https://ni.neatvideo.com/home.

Online. 2015b. [online] available:. https://ni.neatvideo.com/.

Plötz, T., and Roth, S. 2017. Benchmarking denoising algorithms with real photographs. In *CVPR*.

Ren, H.; El-Khamy, M.; and Lee, J. 2018. Dn-resnet: Efficient deep residual network for image denoising. *arXiv preprint arXiv:1810.06766*.

Roth, S., and Black, M. J. 2009. Fields of experts. *International Journal of Computer Vision* 82(2):205.

Santhanam, V.; Morariu, V. I.; and Davis, L. S. 2017. Generalized deep image to image regression. In *CVPR*.

Shi, W.; Caballero, J.; Huszár, F.; Totz, J.; Aitken, A. P.; Bishop, R.; Rueckert, D.; and Wang, Z. 2016. Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network. In *CVPR*.

Shi Guo, Zifei Yan, K. Z. W. Z. L. Z. 2018. Toward convolutional blind denoising of real photographs. In *arXiv preprint arXiv:1807.04686*.

Stanford. 2015. Demosaicking and denoising. https://web.stanford.edu/group/vista/cgi-bin/wiki/index.php/Demosaicking_and_Denoising.

Ulyanov, D.; Vedaldi, A.; and Lempitsky, V. 2017. Deep image prior. *arXiv preprint arXiv:1711.10925*.

Wang, C.; Huang, H.; Han, X.; and Wang, J. 2019. Video inpainting by jointly learning temporal structure and spatial details. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, 5232–5239.

Xu, J.; Ren, D.; Zhang, L.; and Zhang, D. 2016. Patch group based bayesian learning for blind image denoising. In *ACCV*.

Xu, J.; Zhang, L.; Zhang, D.; and Feng, X. 2017. Multi-channel weighted nuclear norm minimization for real color image denoising. In *ICCV*.

Xu, J.; Zhang, L.; and Zhang, D. 2018. A trilateral weighted sparse coding scheme for real-world image denoising. *arXiv preprint arXiv:1807.04364*.

Yair, N., and Michaeli, T. 2018. Multi-scale weighted nuclear norm image restoration. In *CVPR*.

Yu, F., and Koltun, V. 2015. Multi-scale context aggregation by dilated convolutions. *arXiv preprint arXiv:1511.07122*.

Yu, K.; Wang, X.; Dong, C.; Tang, X.; and Loy, C. C. 2019. Path-restore: Learning network path selection for image restoration. *arXiv preprint arXiv:1904.10343*.

Zhang, K.; Zuo, W.; Chen, Y.; Meng, D.; and Zhang, L. 2017a. Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising. *IEEE Transactions on Image Processing* 26(7):3142–3155.

Zhang, K.; Zuo, W.; Gu, S.; and Zhang, L. 2017b. Learning deep cnn denoiser prior for image restoration. In *CVPR*.

Zhang, K.; Zuo, W.; and Zhang, L. 2018. Ffdnet: Toward a fast and flexible solution for cnn based image denoising. *IEEE Transactions on Image Processing*.

Zhao, R.; Lam, K.-M.; and Lun, D. P. 2019. Enhancement of a cnn-based denoiser based on spatial and spectral analysis. In *2019 IEEE International Conference on Image Processing (ICIP)*, 1124–1128. IEEE.

Zhou, Y.; Jiao, J.; Huang, H.; Wang, J.; and Huang, T. 2019. Adaptation strategies for applying awgn-based denoiser to realistic noise. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, 10085–10086.

Zhou, Y.; Liu, D.; and Huang, T. 2018. Survey of face detection on low-quality images. In *2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018)*, 769–773. IEEE.

Zhu, F.; Chen, G.; and Heng, P.-A. 2016. From noise modeling to blind image denoising. In *CVPR*.

Zoran, D., and Weiss, Y. 2011. From learning models of natural image patches to whole image restoration. In *ICCV*.