

C2N: Practical Generative Noise Modeling for Real-World Denoising

Geonwoon Jang* Wooseok Lee* Sanghyun Son Kyoungmu Lee

ASRI, Department of ECE, Seoul National University

{onwoono, adntjr4}@gmail.com, {thstkdgus35, kyoungmu}@snu.ac.kr

Abstract

Learning-based image denoising methods have been bounded to situations where well-aligned noisy and clean images are given, or samples are synthesized from predetermined noise models, e.g., Gaussian. While recent generative noise modeling methods aim to simulate the unknown distribution of real-world noise, several limitations still exist. In a practical scenario, a noise generator should learn to simulate the general and complex noise distribution without using paired noisy and clean images. However, since existing methods are constructed on the unrealistic assumption of real-world noise, they tend to generate implausible patterns and cannot express complicated noise maps. Therefore, we introduce a Clean-to-Noisy image generation framework, namely C2N, to imitate complex real-world noise without using any paired examples. We construct the noise generator in C2N accordingly with each component of real-world noise characteristics to express a wide range of noise accurately. Combined with our C2N, conventional denoising CNNs can be trained to outperform existing unsupervised methods on challenging real-world benchmarks by a large margin.

1. Introduction

Image denoising aims to remove unintended signals from a given noisy observation. The task has been considered as one of the fundamental vision problems and handled by numerous studies [9, 14, 17]. While recent deep convolutional neural networks (CNNs) have achieved promising performance [50, 51, 19, 11, 36], several challenges prevent them from being used for practical applications. A primary limitation of the learning-based approaches is that they are usually data-driven, where training on a specific dataset does not guarantee generalization to a real-world scenarios [53, 19].

The noise from a typical camera pipeline is different to the conventional assumption for ideal noise in several as-

*Authors contributed equally.

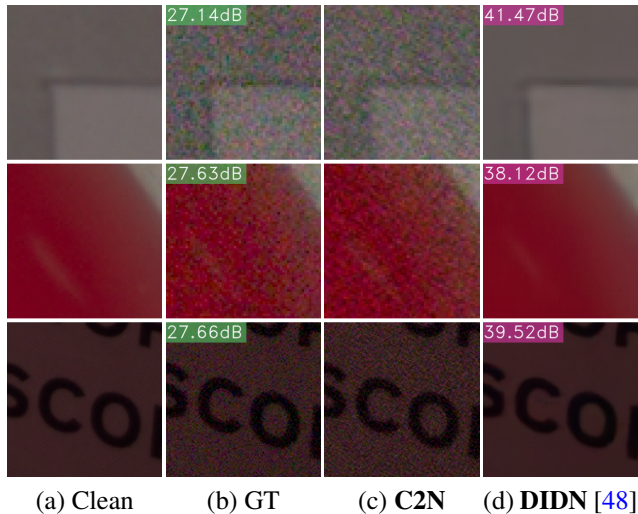


Figure 1: **Examples of generated and denoised image from our proposed method.** (a) Clean image, (b) Ground truth noisy image, (c) Generated noisy image from the proposed C2N, (d) Denoising results of DIDN [48] trained on the images generated by our C2N. Our C2N can accurately imitate the real noise without using paired examples.

pects. For instance, a widely-used Additive White Gaussian Noise (AWGN) formulation assumes that the term is signal-independent [10, 28], while real-world noises are not. Therefore, it is difficult to generalize a denoising algorithm toward real-world images when the model is trained on synthetic examples. As an alternative, few studies have collected well-aligned noisy and clean image pairs in the wild [2, 40] so that the following denoising methods can be trained in a supervised manner. While such an approach is an effective way to deal with real-world noise, it remains challenging to acquire large-scale pairs due to several practical issues. Recent self-supervised approaches [28, 6] deal with the limitations by using noisy samples as the target output, whose supervision would lead a model to estimate the actual value on average. However, they usually leverage some statistical properties of the noise, which are insufficient to represent the real-world cases [2, 40].

On the other hand, generation-based approaches [12, 22] have been proposed to deal with challenging real-world situations. These approaches first train a noise generator [16] on target noisy images to produce pseudo-noisy images paired to other clean images, which are used later to train a denoising model. Following the successful results of the earlier methods [12, 10] for synthetic noise reduction, recent attempts [1, 22] to apply this method to real-world noise have been introduced. Still, in a situation where no paired clean images are given to the target noisy images, a generation-based method that successfully imitates real-world noise has not been suggested.

In this paper, we introduce C2N(Clean to Noisy), a novel generative noise modeling framework trained without any paired data. The C2N can learn a variety of complex noise distributions successfully and generates accurate noisy images from arbitrary clean images. We achieve state-of-the-art performance by training existing denoising models with the generated pairs from our C2N framework. Our contributions is summarized as follows:

- We propose a novel noise generator with explicit modules to express noise terms of according characteristics, enabling the generator to imitate more accurate real-world noise.
- Our C2N framework successfully simulates target noise distribution without any handcrafted formulations or unrealistic assumptions.
- With the data pairs generated by C2N, we train denoising models that outperform state-of-the-art unsupervised methods for denoising real photographs.

2. Related Works

Deep Image Denoising. After the DnCNN [50] model has achieved a significant performance gain over traditional methods [9, 14], CNN-based methods have become mainstream in the image denoising area. The FFDNet [51] model uses given noise level maps to remove AWGN with spatially varying noise levels effectively. Using large models and complex architectures, the performance of denoising models can be improved by extracting rich features from the input noisy image [11, 24]. While such methods successfully erase AWGN with state-of-the-art performances, they still require numerous training pairs that contain the exact target noise distribution. Therefore, the real-world denoising task has remained challenging since the desired noise model is unknown without appropriate training examples.

Deep Denoising of Real-World Noise. If enough amount of training samples are given, it is straightforward to train the methods mentioned to function properly. For such purpose, Anaya and Barbu [4] acquire the Renoir dataset where clean samples are synthesized from a sequence of low-ISO images. However, spatial misalignment and remaining noise in training pairs make it challenging to use the

dataset for practical purpose. The DND [40] dataset post-processes low-ISO images to align their spatial contents and illuminations with high-ISO counterparts. The SIDD [2] dataset captures noisy images under various lighting conditions with five different smartphone cameras. With these real-world noise datasets, denoising models with various attention modules [25, 5], multi-scale resizing in features [42, 35, 48], or use of self-similarity in images [30, 34, 46] manage to remove complex real noises.

Nevertheless, it is difficult to collect large-scale real-world dataset for our specific purpose. To remove noise even when the accurate distribution of it is not given (e.g. blind denoising), the CBDNet [19] includes a part that performs noise level estimation based on heteroscedastic Gaussian noise model [15]. It is intended to also operate on real noisy images, receiving indirect supervision from training the denoiser. The Path-Restore [47] dynamically selects an appropriate restoration path for each region of an input image. The self-supervised denoising methods [28, 6] use only individual noisy images for training and estimates a pixel value of its input noisy image itself, where the value of that location is masked-out as 'blind spot'. Since these methods require the noise to satisfy strong statistical assumptions, [8, 29] modify these according to the prior knowledge of noisy images. Recently, the AINDNet [26] apply adaptive instance normalizing method to deal with varying noise levels. And the Noise2Blur [33] performs an additional procedure to preserve image details after training a model with blurred labels.

Noise Generation-Based Denoising. The generation-based methods [12, 10, 1, 22] usually adopt a two-stage pipeline for the denoising problem rather than using a single model. First, a noise generator is learned in an unsupervised fashion [16] to simulate the distribution of given real noisy examples so that any clean images can be mapped to pseudo-noisy data. A denoising model can then be trained in a straightforward manner using the synthesized input and target pairs. The GCBD [12] is the first generation-based approach for deep blind image denoising. Later, the GAN2GAN [10] method leverages better noisy-patch-extraction, generating more realistic noisy samples to train the following denoising model in an N2N [31] manner. While the methods above are limited to show promising results for signal-independent and synthetic noise only, the Noise Flow [1] proposes a formulation to imitate challenging real-world noise. By leveraging the normalizing flow formed of invertible transforms, it can precisely learn the shift of distribution between synthetic and desired noise maps. Although the model successfully imitates in-camera noise occurring pipeline, true noisy and clean image pairs are required to get the correct noise distribution, which is not practical. The DANet [49] also learns the mapping of denoiser and generator by comparing them with the true

noisy-to-clean joint distribution, but it still needs paired images to learn such mapping. Recently, the UIDNet [22] model handles the unpaired generative noise modeling with the sharpening technique for better noise separation. Furthermore, the NTGAN [52] method demonstrates that noise maps synthesized with a given camera response function (CRF) can be used for the following denoising network. Unlike the existing generation-based denoising methods, we introduce a novel noise generator trained without any paired data or heuristic methods.

3. Method

3.1. Complexity of Real-World Noise

To generate realistic noise with CNNs, it is necessary to understand the properties and statistical behavior of real-world noise. Due to several physical limitations, the noise occurs from various sources, including electronic sensors, in-camera amplifiers, photon noise, quantization, and compression artifacts [21]. Combining all these factors, the pixel-wise noise term n is mixed with an underlying clean signal x , resulting the noisy observation y as follows:

$$y = x + n. \quad (1)$$

In conventional deep denoising methods [50, 51], the noise term n is usually simplified as an ideal Additive White Gaussian Noise (AWGN), i.e., $n \sim \mathcal{N}(0, \sigma^2)$, where σ denotes the standard deviation. On the other hand, the photon noise is signal-dependent, where Poisson distribution can be used to simulate the case and be approximated as a Gaussian distribution with signal-dependent variance [15]. The heteroscedastic Gaussian noise is defined as follows:

$$n \sim \mathcal{N}(0, \sigma_s^2 x + \sigma_c^2), \quad (2)$$

where σ_s and σ_c are hyperparameters for signal-dependent and signal-independent term. While the noise model in (2) can provide a proper approximation of the realistic noise [19, 52] to some extent, several studies [21, 45] have demonstrated that the real-world cases appear to be much more complicated. In addition, the physical limitations of in-camera electronic devices and in-camera compression pipeline make the noise term to exhibit random spatial pattern [7]. Such property leads the noise term n and its local neighbors to be spatially correlated, making the precise modeling more challenging.

To deal with the problem without using paired data, previous approaches [19, 52] construct noise maps using a synthetic noise model and known camera response function (CRF). However, such handcrafted features prevent those methods from being generalized toward realistic configurations. Therefore, we adopt a learning-based method, namely C2N, to simulate the real-world noise rather than

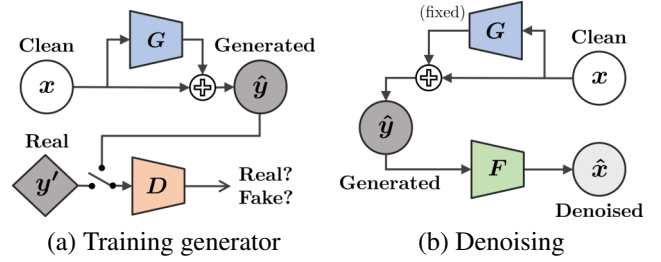


Figure 2: **Our two-step pipeline for real-world denoising.** (a) Our method first learns to generate samples from target noise distribution, with a noise generator G . We used clean image x and noise image y' to be unpaired. (b) Secondly, using the generated pairs, we train a denoising model F .

using some handcrafted formulations. Our framework fully utilizes the advantage of unsupervised learning to simulate the comprehensive real-world noise, with novel design components and objective terms.

3.2. Learning to Generate Pseudo-Noise Images

A denoising network F aims to reconstruct the underlying clean signal x from a given noisy observation y in (1). When enough training pairs are available, the model can be trained in a supervised manner to estimate the clean signal. However, in real-world scenarios, it is challenging to acquire *ideal* clean images well-aligned to the training noisy images, even with complicated post-processing [2]. Hence, our C2N framework first trains a generator to simulate the target noise distribution. Then, the following denoising model F can be learned on the generated noisy examples. Figure 2 shows the two-step pipeline of our method.

Our noise generator network G is designed to synthesize a realistic noise map \hat{n} for a given clean image x to produce the pseudo-noisy image \hat{y} as follows:

$$\hat{y} = x + \hat{n} = x + G(x, r), \quad (3)$$

where r is a random vector to reflect the stochastic behavior of noise according to the conditions of each scene. We sample 32-dim random vector from $\mathcal{N}(0, 1^2)$ and spatially replicate through all pixel positions of x , similar to the GAN applications [38, 41].

Simultaneously, we train a discriminator network D to distinguish whether a given noisy image is synthesized from our generator G or sampled from the real-world dataset. The two networks G and D can be optimized in an adversarial way [16], using the Wasserstein distance [18] as follows:

$$\begin{aligned} \mathcal{L}_{\text{adv}}(D, G) = & \mathbb{E}_{y' \sim P_N} [D(y')] \\ & - \mathbb{E}_{x \sim P_C, r \sim P_r} [1 - D(x + G(x, r))] \\ & + \lambda \mathbb{E}_{x_\delta \sim P_\delta} [(\|\nabla_{x_\delta} D(x_\delta)\|_2 - 1)^2], \end{aligned} \quad (4)$$

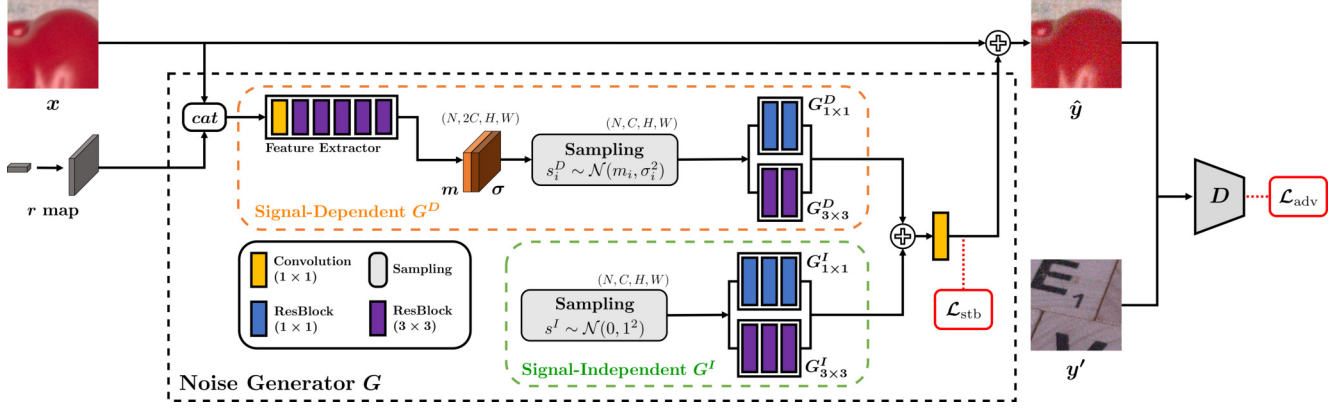


Figure 3: **Overview of C2N framework.** The noise generator G architecture of our C2N and its components are shown. r map denotes the spatially replicated input random vector r . Initial noise maps s^I and s^D are sampled in the signal-independent part G^I and signal-dependent part G^D of the generator, respectively.

where the term is minimized with respect to G and maximized with respect to D . P_N and P_C denote the distribution of the real-world noisy and clean images, respectively. The real noisy image y' is sampled from P_N . Prime notation on y denotes that we use noisy image unpaired to clean image x . P_r is the distribution of random vector r . For stable learning, we adopt the gradient penalty [18], which is weighted by a hyperparameter $\lambda = 10$. The term $x_\delta \sim P_\delta$ is one of the internal dividing points between generated and real images.

The significant advantage of the proposed C2N framework is that generator in the C2N can synthesize realistic noise without adopting handcrafted features. However, the generated noise \hat{n} may bias the image color and negatively affect the overall framework if our C2N is trained without any constraints. To deal with the case, we additionally define a stabilizing loss term \mathcal{L}_{stb} , which is defined as follows:

$$\mathcal{L}_{\text{stb}} = \frac{1}{N} \sum_c \left\| \sum_{i \in B} \hat{n}_{i,c} \right\|_1, \quad (5)$$

where N denotes the number of pixel i in mini-batch B and c is index of each color channels. By minimizing the stabilizing loss, the channel-wise average of the generated noise approaches zero, which prevents the color-shifting problem. Since we do not take a mean over a single sample or local area, the generated noise may have varying nonzero local means depending on the underlying signal. Combining our two loss terms (4) and (5), we optimize the total loss $\mathcal{L}_G = \mathcal{L}_{\text{adv}} + w_{\text{stb}} \mathcal{L}_{\text{stb}}$, where $w_{\text{stb}} = 0.01$.

3.3. C2N Architecture

The previous generation-based approaches have limited their scopes to signal-independent [12] or spatially uncorrelated [1] noise terms. On the other hand, we construct a new generator architecture to represent diverse and com-

plex noise distributions discussed in Section 3.1. We gradually implement several design components to express more general properties of real-world noise. Figure 3 shows the overall architecture of our C2N framework.

Signal-Independent Pixel-Wise Transform. The noise generation process can be formulated as a nonlinear mapping from an initial random noise map to the desired noise map. To generate signal-independent noise, we sample the initial random noise map s^I from the standard normal distribution $\mathcal{N}(0, 1^2)$. We construct a signal-independent pixel-wise noise transformation module $G^I_{1 \times 1}$ to simulate spatially i.i.d. noise. $G^I_{1 \times 1}$ module consists of 3 modified residual blocks [20] for low-level vision problems [32] with 1×1 convolutional layers without batch-normalization.

Signal-Dependent Pixel-Wise Transform. To express signal-dependent noise, our noise generator should extract useful features from the clean input image. However, it is not desirable that the generated noise map \hat{n} is deterministically generated from the given clean image. To effectively represent the signal-dependent noise term, we sample initial noise map s^D from position-wise normal distribution which has mean m_i and standard deviation σ_i at each position i from convolutional features. Input clean image and random vector r are transformed to convolutional features through the feature extractor. Feature extractor consists of 5 residual blocks after 1×1 convolutional layer without using any pre-trained model. We then sample the initial noise map s^D_i at each position i , and apply another pixel-wise noise transforms $G^D_{1 \times 1}$ which consists of 2 residual blocks with 1×1 convolutional layer. For the sampling of s^D_i , we used reparameterization trick [43] to preserve gradients of the parameters, so that the feature extractor can be jointly trained with the entire C2N framework.

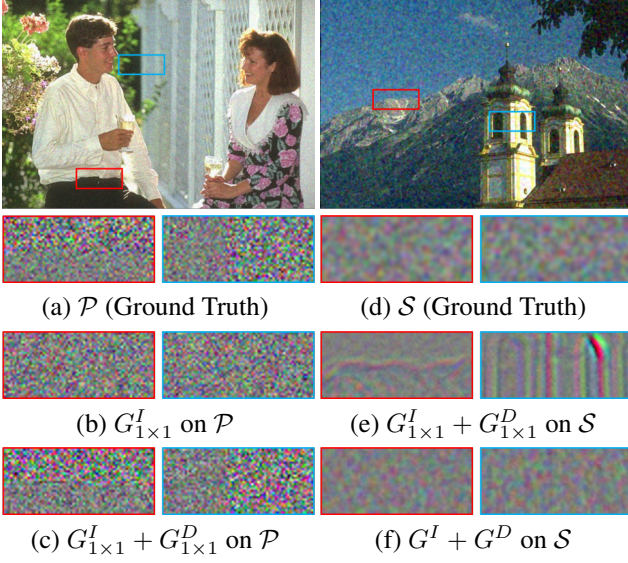


Figure 4: **Generated samples from model ablation study on synthetic noise.** (a, d) Synthetic ground-truth noise map of Poisson \mathcal{P} and spatially correlated Gaussian noise \mathcal{S} . (b, c) Generated residual noise map of the C2N variant trained on the Poisson noise. (e, f) Synthesized residual noise on the spatially correlated Gaussian noise. We amplify the noise maps in (a-c) by 4 times for clearer comparison.

Spatially Correlated Transforms. A number of existing manual noise models [15] and denoising methods assume spatially uncorrelated noise. On the other hand, C2N handles color sRGB image containing various conversion and compression degradations from in-camera post-processing [7] in end-to-end manner. To achieve this, We add transforms of 3×3 convolution like $G_{1 \times 1}^I$ and $G_{1 \times 1}^D$, which are denoted as $G_{3 \times 3}^I$ and $G_{3 \times 3}^D$. These 3×3 convolutions involve capability for express spatially correlated noise term. $G_{3 \times 3}^I$ and $G_{3 \times 3}^D$ produce features for the *spatially correlated signal-independent noise* and the *spatially correlated signal-dependent noise*, respectively.

Lastly, we add all the noise features transformed through the $G_{1 \times 1}^I$, $G_{3 \times 3}^I$, $G_{1 \times 1}^D$, $G_{3 \times 3}^D$ into one, resulting in a noise feature that integrates the characteristics of noise expressed by each module. Then we take 1×1 convolution to the merged feature map to reduce the dimension to the color space, and also to perform a non-linear mapping from the integrated feature to the final noise map. We set the number of layer channels C as 64, and all the intermediate features from s^I and s^D have the same number of channels.

3.4. Learning to Denoise with the Generated Pairs

With the C2N, it is straightforward to optimize the following denoising network F . We first generate pseudo-noisy images \hat{y} from the clean examples x and use the pairs to train a denoising model in a supervised manner [50, 51].

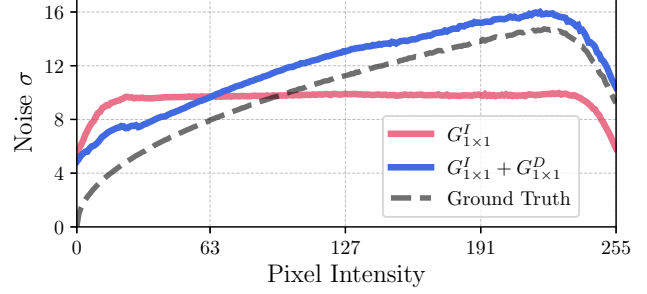


Figure 5: **Comparison between generated noise and ground truth Poisson noise.** We illustrate pixel noise level with respect to its underlying ground truth signal intensity. Noise level is measured on the CBSD68 dataset. Distortions around boundary values are due to the clipping of the over-exposed and under-illuminated pixels [2].

Similar to the previous deep denoising methods [48, 25], we minimize the L_1 reconstruction loss which is defined as follows:

$$\mathcal{L}_{\text{rec}} = \frac{1}{m} \sum_{k=1}^m \|F(\hat{y}) - x\|_1, \quad (6)$$

where \hat{y} is pseudo noisy image generated by $\hat{y} = x + G(x, r)$, on x sampled from clean images, and k is index of each images in a mini-batch B of size m .

The major advantages of our approach is that our framework is independent to the selection of following denoising architecture. Previous attempts like [10, 22] train their noise generator and denoising model jointly. Since the C2N model doesn't get any supervision from the reconstruction loss \mathcal{L}_{rec} , the generated images are not specialized for certain denoising model.

4. Experiment

4.1. Experimental Setup

Dataset. To train and evaluate our method on synthetic noise, we use clean color images from the BSD500 [37] dataset, which consists of 432 training images and 68 test samples. To train our C2N and denoising model on challenging real-world noise, we leverage the DND [40] and the SIDD [2] dataset captured under realistic environments. The training split of the SIDD dataset, i.e., SIDD Medium, contains 320 noisy and clean image pairs and the DND dataset contains 50 noisy images. Since it does not provide any ground truth image for learning on DND, we combine the clean images of the SIDD and noisy images in the DND. To train our C2N framework, we crop 36,000 patches of size 96×96 from the noisy and clean image sets and randomly sample them to gather training mini-batches. For training denoising network, we use only clean patches of the same size and number. We evaluate the denoising models followed by the C2N with the DND and the SIDD benchmarks.

Training Data	Test Noise Level		
	$\sigma = 15$	$\sigma = 25$	$\sigma = 50$
Synthetic	$\sigma = 15$	33.48	27.24
	$\sigma = 25$	31.39	30.68
	$\sigma = 50$	27.82	27.81
C2N	32.96	30.51	27.09

Table 1: **Denoising performance on synthetic AWGN.** PSNR(dB) is calculated on the CBSD68 dataset. We note that the C2N models are trained for each noise level independently.

	AWGN	UIDNet	C2N (Ours)
KL-divergence	0.1746	0.4417	0.1638

Table 2: **KL-divergence between generated noise map and ground truth noise map.** The values are calculated using the SIDD validation set. We refer the readers to the supplementary materials for more details on the metric.

$G_{1 \times 1}^I$	$G_{1 \times 1}^D$	$G_{3 \times 3}^I$	$G_{3 \times 3}^D$	\mathcal{L}_{stb}	PSNR(dB)
✓	✓	✓	✓		9.81
✓				✓	28.54
	✓			✓	30.46
✓	✓			✓	31.74
✓		✓		✓	32.19
	✓		✓	✓	32.21
✓	✓	✓	✓	✓	34.08

Table 3: **Model ablation study on SIDD validation set.** The notation of each modules follows the Figure 3. Note that feature extractor module exists when any signal-dependent module is on.

Implementation Details and Optimization. To optimize our C2N framework, we augment all training samples by randomly flipping and rotating them by 90° to construct a mini-batch of size 36. The Adam [27] optimizer is used with an initial learning rate of 10^{-4} . The learning rate is multiplied by a factor of 0.8 for every 3 epochs, where a single C2N model is trained over 36 epochs. Denoising models are optimized using only the generated training pairs from trained C2N generator with mini-batch of size 16. Same with C2N, the Adam [27] optimizer is used with an initial learning rate of 10^{-4} and learning rate is halved for every 4 epochs. The training runs for total 16 epochs. We select the CDnCNN-B [50] architecture for denoising as our baseline unless otherwise stated.

4.2. Model Analysis of C2N on Synthetic Noise

To demonstrate the validity and effectiveness of our C2N framework, we first analyze how each of the design component supports the C2N architecture in handling the synthetic noise of various properties. Figure 4 illustrates how each modules in our C2N model can imitate two synthetic noise

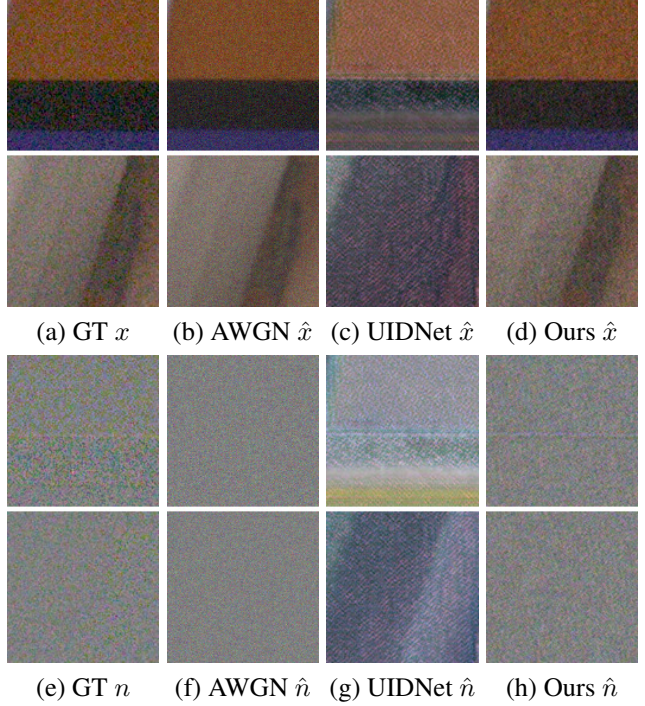


Figure 6: **Visual comparison of the generated noisy samples on SIDD dataset.** (a) A ground truth noisy image (b-d) Generated noisy image of AWGN, UIDNet [22] and Ours respectively. (e-h) are the residual noise maps of each (a-d).

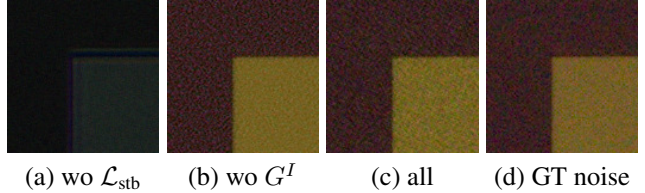


Figure 7: **Generated samples from model ablation study on real-world noise.** (a) without stabilizing loss, (b) without independent transforms, (c) with all modules and stabilizing loss, (d) Ground truth noisy image.

models, \mathcal{P} and \mathcal{S} . \mathcal{P} stands for signal-dependent Poisson noise, and \mathcal{S} stands for spatially correlated Gaussian noise of $\sigma = 50$. To implement spatial correlation between local neighbors, we apply a 9×9 Gaussian filter to the noise map, similar to [39]. Then, we train C2N variants with different existence of each module, e.g., $G_{1 \times 1}^I$, $G_{1 \times 1}^D$, G^I , and G^D , on BSD images corrupted by \mathcal{P} and \mathcal{S} . We note that the notation G^I and G^D refers to the combination of ($G_{1 \times 1}^I$, $G_{3 \times 3}^I$) and ($G_{1 \times 1}^D$, $G_{3 \times 3}^D$), respectively. More details are described in our supplementary material.

The model only with signal-independent module, i.e., $G_{1 \times 1}^I$ that equally generates noise values at all locations, is not proper to synthesize signal-dependent noise. Therefore, the output noise in Figure 4b does not differ in dark

	Method	SIDD		DND	
		PSNR(dB)	SSIM	PSNR(dB)	SSIM
Non-learning based	BM3D [14]	25.65	0.685	34.51	0.851
	WNNM [17]	25.78	0.809	34.67	0.865
	K-SVD [3]	26.88	0.842	36.49	0.898
	EPLL [23]	27.11	0.870	33.51	0.824
Supervised	TNRD [13]	24.73	0.643	33.65	0.831
	DnCNN [50]	23.66	0.583	32.43	0.790
	DnCNN ⁺ [50]	35.13	0.896	37.89	0.932
	CBDNet [19]	33.28	0.868	38.05	0.942
	AINDNet(R)* [26]	38.84	0.951	39.34	0.952
	DIDN [48]	39.82	0.973	-	-
	DANet [49]	39.43	0.956	39.58	0.955
Self-supervised	N2V [28]	27.68	0.668	-	-
Generation-based	GCBD [12]	-	-	35.58	0.922
	UIDNet-NS [22]	31.34	0.856	-	-
	UIDNet [22]	32.48	0.897	-	-
	C2N + DnCNN (Ours)	33.76	0.901	36.08	0.903
	C2N + DnCNN* (Ours)	34.00	0.907	36.32	0.908
	C2N + DIDN (Ours)	35.02	0.932	36.12	0.882
	C2N + DIDN* (Ours)	35.35	0.937	36.38	0.887
	C2N(SD) + DIDN* (Ours)	-	-	37.28	0.924

Table 4: **Quantitative evaluation on the SIDD and DND benchmark.** We adopt the two-stage pipeline which is denoted by ‘C2N + Denoiser’. * denotes the method with self-ensemble [44, 32] strategy. (SD) denotes that the C2N generator is first trained on the SIDD and then fine-tuned on the DND. Results of DnCNN and DnCNN⁺ are from the model trained on synthetic noise and the model trained by ourselves on SIDD images, respectively.

and bright areas compared to the ground-truth distribution in Figure 4a. By explicitly considering the signal-dependent component with the $G_{1 \times 1}^D$ module, our C2N can express such behavior of the Poisson noise as shown in Figure 4c. Figure 5 further validates that our generator can synthesize signal-dependent noise, where standard deviations of the noise intensities are correlated with pixel values. Figure 4e illustrates that a model with only 1×1 convolutions, i.e., $G_{1 \times 1}^I + G_{1 \times 1}^D$, cannot imitate spatially correlated noise and generates undesirable structures. By including modules with 3×3 convolutions, i.e., $G_{3 \times 3}^I$ and $G_{3 \times 3}^D$, the proposed C2N can synthesize the desired noise distribution without unleashing artifacts, as shown in Figure 4f.

We also evaluate our method on synthetic AWGN and compare the performance with supervised denoising models in Table 1. The supervised model requires pairs of clean and noisy examples that correspond to the target noise level. Otherwise, the denoising performances are degraded due to a mismatch in training and test distribution. However, such an assumption is less practical in the real-world scenario as we may not know the exact noise levels of our target images. The primary advantage of the proposed C2N framework is that we leverage unpaired noisy and clean images, and the exact noise level is not required to develop the following denoising method. Our unsupervised generation-based

method achieves comparable performance to the supervised denoising model on a synthetic Gaussian dataset.

4.3. Model Analysis of C2N on Real-World Noise

We validate the effectiveness of each modules in our C2N framework under the real-world degradations. Figure 6 shows that the generator in our C2N can generate realistic noise maps, whereas the existing noise generator tends to produce a particular texture. Also, our C2N achieves the lowest KL-divergence between the generated and ground truth noise maps, among the compared methods, as show in Table 2. The stochastic behavior and explicit transforms of the C2N model help its output to maintain the characteristics of noise, solving the difficulties of a regular CNN functioning as a noise generator. Note that for UIDNet experiment in Figure 6 and Table 2, we train model with officially released code.

Also, Table 3 and Figure 7 show that each component of our C2N model is essential for real-world noise modeling. Since our method learns the noise distribution without any accurate noise maps or a heuristic technique to stabilize the learning, the stabilizing loss \mathcal{L}_{stb} plays an important role. By comparing rows with only 1×1 module to rows combined with 3×3 module of Table 3, we can again verify the necessity of G^I , G^D and 3×3 convolutional transforms. In

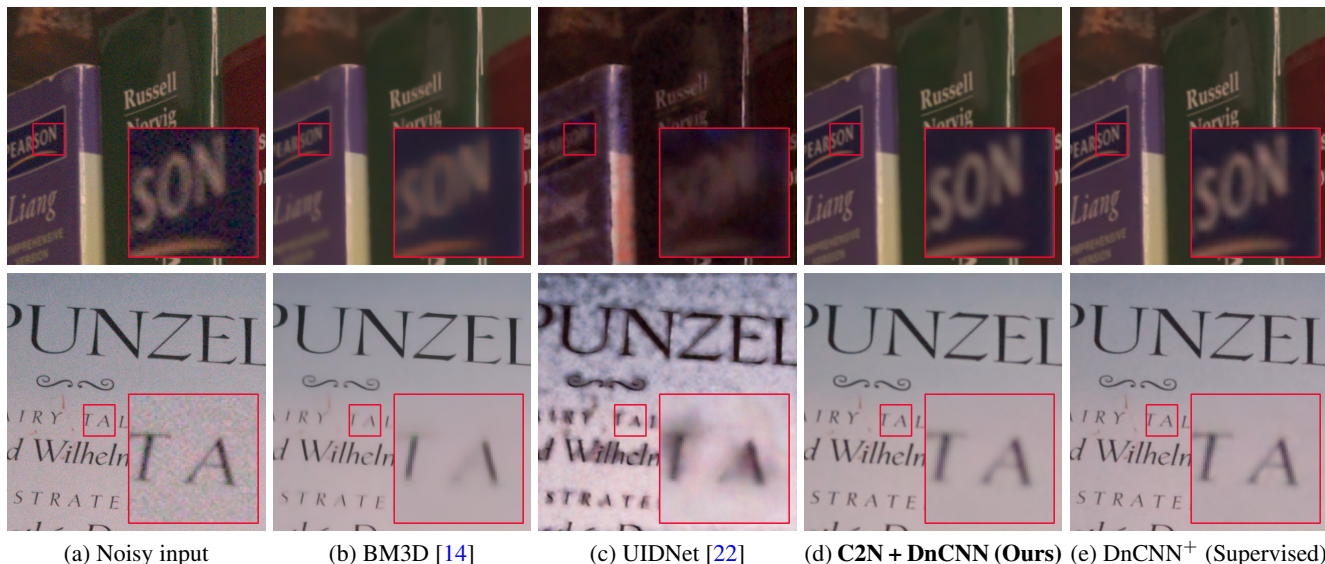


Figure 8: **Comparison between denoising result examples of the SIDD dataset.** (c) We use publicly released UIDNet to get the result images. (e) The DnCNN model is trained in a supervised manner, using clean-noisy pairs in the SIDD dataset.

addition, using signal-independent and dependent module together gain higher PSNR result than others. Result shows that each single module is not enough for generating complex real-world noise. For example, the model without G^I at the 6th row of the Table 3 produces samples that differ in noise distribution, as shown in Figure 7. From the ablation study, we confirm that the entire elements of the C2N together can well simulate challenging real-world noise.

4.4. Results of Real-World Denoising

Table 4 shows the performance of our method evaluated on the DND and the SIDD benchmarks of real-world denoising in sRGB space. We note that our method outperforms existing unsupervised methods by a large margin. Especially, our method with C2N and DnCNN denoiser shows better results on the SIDD benchmark than the UIDNet which use similar denoising network with DnCNN model. We also use much fewer patch data than the 520,965 noisy and clean image patches of 64×64 size used by UIDNet. We further improve the performance of our method by using the DIDN [48] as its denoising model, which has larger capacity of $\sim 217M$ parameters compared to that of $\sim 0.67M$ parameters of the DnCNN. It shows that we can train arbitrary denoising models with C2N, unlike UIDNet with a fixed denoising backbone.

In the DND benchmark, our method also outperforms GCBD, another previous generation-based method. Several existing methods [19, 26] use rich training images of the SIDD to get high performance on the DND benchmark, even though the DND and SIDD datasets are not matching datasets. Similarly, we also train our C2N generator on the SIDD dataset and fine-tune trained model on the DND

dataset to obtain more improved performance than directly training on the DND, as shown in the last row of Table 4.

Figure 8 shows denoised result examples on the SIDD dataset. Along with the quantitative results, our method shows better noise removal and detail-preserving results than the compared unsupervised methods, with quality comparable to that of the supervised model.

5. Conclusion

In this paper, we propose a C2N framework for practical real-world denoising which includes our novel noise generator. By explicitly designing components of the generator considering signal-dependency and spatial correlation of real-world noise property, it successfully learn to simulate the noise distribution of the noisy images in situation of unpaired setting. We show that each module of the model can generate noise with corresponding characteristics through the experiments on several synthetic noise and real-world noise. Using the generated noisy and clean pairs from our generator, we train denoisers to outperform the existing methods without use of actual data pairs. We believe our method can be a key to solve the challenging points of practical real-world denoising.

Acknowledgement

This work was supported in part by an IITP grant funded by the Korean government [No. 2021-0-01343, Artificial Intelligence Graduate School Program (Seoul National University)], and in part by AIRS Company in Hyundai Motor Company and Kia Corporation through HMC/KIA-SNU AI Consortium Fund.

References

- [1] Abdelrahman Abdelhamed, Marcus A Brubaker, and Michael S Brown. Noise flow: Noise modeling with conditional normalizing flows. In *ICCV*, 2019. 2, 4
- [2] Abdelrahman Abdelhamed, Stephen Lin, and Michael S Brown. A high-quality denoising dataset for smartphone cameras. In *CVPR*, 2018. 1, 2, 3, 5
- [3] Michal Aharon, Michael Elad, and Alfred Bruckstein. K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation. *IEEE Transactions on Signal Processing*, 54(11):4311–4322, 2006. 7
- [4] Josue Anaya and Adrian Barbu. RENOIR—a dataset for real low-light image noise reduction. *Journal of Visual Communication and Image Representation*, 51:144–154, 2018. 2
- [5] Saeed Anwar and Nick Barnes. Real image denoising with feature attention. In *ICCV*, 2019. 2
- [6] Joshua Batson and Loic Royer. Noise2Self: Blind denoising by self-supervision. In *ICML*, 2019. 1, 2
- [7] Alan C Bovik. *The essential guide to image processing*. Academic Press, 2009. 3, 5
- [8] Coleman Broaddus, Alexander Krull, Martin Weigert, Uwe Schmidt, and Gene Myers. Removing structured noise with self-supervised blind-spot networks. In *ISBI*, 2020. 2
- [9] Antoni Buades, Bartomeu Coll, and J-M Morel. A non-local algorithm for image denoising. In *CVPR*, 2005. 1, 2
- [10] Sungmin Cha, Taeon Park, Byeongjoon Kim, Jongduk Baek, and Taesup Moon. GAN2GAN: Generative noise learning for blind denoising with single noisy images. In *ICLR*, 2021. 1, 2, 5
- [11] Chang Chen, Zhiwei Xiong, Xinmei Tian, and Feng Wu. Deep boosting for image denoising. In *ECCV*, 2018. 1, 2
- [12] Jingwen Chen, Jiawei Chen, Hongyang Chao, and Ming Yang. Image blind denoising with generative adversarial network based noise modeling. In *CVPR*, 2018. 2, 4, 7
- [13] Yunjin Chen and Thomas Pock. Trainable nonlinear reaction diffusion: A flexible framework for fast and effective image restoration. *TPAMI*, 39(6):1256–1272, 2016. 7
- [14] Kostadin Dabov, Alessandro Foi, Vladimir Katkovnik, and Karen Egiazarian. Image denoising by sparse 3-d transform-domain collaborative filtering. *TIP*, 16(8):2080–2095, 2007. 1, 2, 7, 8
- [15] Alessandro Foi, Mejdi Trimeche, Vladimir Katkovnik, and Karen Egiazarian. Practical poissonian-gaussian noise modeling and fitting for single-image raw-data. *TIP*, 17(10):1737–1754, 2008. 2, 3, 5
- [16] Ian J Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial networks. *NIPS*, 2014. 2, 3
- [17] Shuhang Gu, Lei Zhang, Wangmeng Zuo, and Xiangchu Feng. Weighted nuclear norm minimization with application to image denoising. In *CVPR*, 2014. 1, 7
- [18] Ishaan Gulrajani, Faruk Ahmed, Martin Arjovsky, Vincent Dumoulin, and Aaron C Courville. Improved training of wasserstein gans. In *NIPS*, 2017. 3, 4
- [19] Shi Guo, Zifei Yan, Kai Zhang, Wangmeng Zuo, and Lei Zhang. Toward convolutional blind denoising of real photographs. In *CVPR*, 2019. 1, 2, 3, 7, 8
- [20] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *CVPR*, 2016. 4
- [21] Gerald C Holst. Ccd arrays, cameras, and displays. 1998. 3
- [22] Zhiwei Hong, Xiaocheng Fan, Tao Jiang, and Jianxing Feng. End-to-end unpaired image denoising with conditional adversarial networks. In *AAAI*, 2020. 2, 3, 5, 6, 7, 8
- [23] Samuel Hurault, Thibaud Ehret, and Pablo Arias. Epil: an image denoising method using a gaussian mixture model learned on a large set of patches. *Image Processing On Line*, 8:465–489, 2018. 7
- [24] Xixi Jia, Sanyang Liu, Xiangchu Feng, and Lei Zhang. FOC-Net: A fractional optimal control network for image denoising. In *CVPR*, 2019. 2
- [25] Dong-Wook Kim, Jae Ryun Chung, and Seung-Won Jung. GRDN: Grouped residual dense network for real image denoising and gan-based real-world noise modeling. In *CVPRW*, 2019. 2, 5
- [26] Yoonsik Kim, Jae Woong Soh, Gu Yong Park, and Nam Ik Cho. Transfer learning from synthetic to real-noise denoising with adaptive instance normalization. In *CVPR*, 2020. 2, 7, 8
- [27] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014. 6
- [28] Alexander Krull, Tim-Oliver Buchholz, and Florian Jug. Noise2Void-learning denoising from single noisy images. In *CVPR*, 2019. 1, 2, 7
- [29] Samuli Laine, Tero Karras, Jaakko Lehtinen, and Timo Aila. High-quality self-supervised deep image denoising. In *NIPS*, 2019. 2
- [30] Stamatis Lefkimmiatis. Non-local color image denoising with convolutional neural networks. In *CVPR*, 2017. 2
- [31] Jaakko Lehtinen, Jacob Munkberg, Jon Hasselgren, Samuli Laine, Tero Karras, Miika Aittala, and Timo Aila. Noise2Noise: Learning image restoration without clean data. In *ICML*, 2018. 2
- [32] Bee Lim, Sanghyun Son, Heewon Kim, Seungjun Nah, and Kyoung Mu Lee. Enhanced deep residual networks for single image super-resolution. In *CVPRW*, 2017. 4, 7
- [33] Huangxing Lin, Weihong Zeng, Xinghao Ding, Xueyang Fu, Yue Huang, and John Paisley. Noise2Blur: Online noise extraction and denoising. *arXiv preprint arXiv:1912.01158*, 2019. 2
- [34] Ding Liu, Bihan Wen, Yuchen Fan, Chen Change Loy, and Thomas S Huang. Non-local recurrent network for image restoration. In *NIPS*, 2018. 2
- [35] Pengju Liu, Hongzhi Zhang, Kai Zhang, Liang Lin, and Wangmeng Zuo. Multi-level wavelet-cnn for image restoration. In *CVPRW*, 2018. 2
- [36] Yang Liu, Saeed Anwar, Liang Zheng, and Qi Tian. GradNet image denoising. In *CVPRW*, 2020. 1
- [37] David Martin, Charless Fowlkes, Doron Tal, and Jitendra Malik. A database of human segmented natural images

- and its application to evaluating segmentation algorithms and measuring ecological statistics. In *ICCV*, 2001. [5](#)
- [38] Mehdi Mirza and Simon Osindero. Conditional generative adversarial nets. *arXiv preprint arXiv:1411.1784*, 2014. [3](#)
 - [39] Nick Moran, Dan Schmidt, Yu Zhong, and Patrick Coady. Noisier2Noise: Learning to denoise from unpaired noisy data. In *CVPR*, 2020. [6](#)
 - [40] Tobias Plotz and Stefan Roth. Benchmarking denoising algorithms with real photographs. In *CVPR*, 2017. [1](#), [2](#), [5](#)
 - [41] Scott Reed, Zeynep Akata, Xincheng Yan, Lajanugen Logeswaran, Bernt Schiele, and Honglak Lee. Generative adversarial text to image synthesis. In *ICML*, 2016. [3](#)
 - [42] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-Net: Convolutional networks for biomedical image segmentation. In *MICCAI*, 2015. [2](#)
 - [43] John Schulman, Nicolas Heess, Theophane Weber, and Pieter Abbeel. Gradient estimation using stochastic computation graphs. In *NIPS*, 2015. [4](#)
 - [44] Radu Timofte, Rasmus Rothe, and Luc Van Gool. Seven ways to improve example-based single image super resolution. In *CVPR*, pages 1865–1873, 2016. [7](#)
 - [45] Kaixuan Wei, Ying Fu, Jiaolong Yang, and Hua Huang. A physics-based noise formation model for extreme low-light raw denoising. In *CVPR*, 2020. [3](#)
 - [46] Zhihao Xia and Ayan Chakrabarti. Identifying recurring patterns with deep neural networks for natural image denoising. In *WACV*, 2020. [2](#)
 - [47] Ke Yu, Xintao Wang, Chao Dong, Xiaoou Tang, and Chen Change Loy. Path-Restore: Learning network path selection for image restoration. *arXiv preprint arXiv:1904.10343*, 2019. [2](#)
 - [48] Songhyun Yu, Bumjun Park, and Jechang Jeong. Deep iterative down-up cnn for image denoising. In *CVPRW*, 2019. [1](#), [2](#), [5](#), [7](#), [8](#)
 - [49] Zongsheng Yue, Qian Zhao, Lei Zhang, and Deyu Meng. Dual adversarial network: Toward real-world noise removal and noise generation. In *ECCV*, 2020. [2](#), [7](#)
 - [50] Kai Zhang, Wangmeng Zuo, Yunjin Chen, Deyu Meng, and Lei Zhang. Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising. *TIP*, 26(7):3142–3155, 2017. [1](#), [2](#), [3](#), [5](#), [6](#), [7](#)
 - [51] Kai Zhang, Wangmeng Zuo, and Lei Zhang. FFDnet: Toward a fast and flexible solution for cnn-based image denoising. *TIP*, 27(9):4608–4622, 2018. [1](#), [2](#), [3](#), [5](#)
 - [52] Rui Zhao, Daniel PK Lun, and Kin-Man Lam. NTGAN: Learning blind image denoising without clean reference. In *BMVC*, 2020. [3](#)
 - [53] Yuqian Zhou, Jianbo Jiao, Haibin Huang, Yang Wang, Jue Wang, Honghui Shi, and Thomas Huang. When AWGN-based denoiser meets real noises. In *AAAI*, 2020. [1](#)