End-to-End Unpaired Image Denoising with Conditional Adversarial Networks

Zhiwei Hong,[1] **Xiaochen Fan,**[2] **Tao Jiang,**[3,1*] **Jianxing Feng**[2*]
[1]Tsinghua University, [2]Haohua Technology Co., Ltd
[3]University of California, Riverside
hzw17@mails.tsinghua.edu.cn, {fanxiaochen, fengjianxing}@harmon.health,
jiang@cs.ucr.edu

Abstract

Image denoising is a classic low level vision problem that attempts to recover a noise-free image from a noisy observation. Recent advances in deep neural networks have outperformed traditional prior based methods for image denoising. However, the existing methods either require paired noisy and clean images for training or impose certain assumptions on the noise distribution and data types. In this paper, we present an end-to-end unpaired image denoising framework (UIDNet) that denoises images with only unpaired clean and noisy training images. The critical component of our model is a noise learning module based on a conditional Generative Adversarial Network (cGAN). The model learns the noise distribution from the input noisy images and uses it to transform the input clean images to noisy ones without any assumption on the noise distribution and data types. This process results in pairs of clean and pseudo-noisy images. Such pairs are then used to train another denoising network similar to the existing denoising methods based on paired images. The noise learning and denoising components are integrated together so that they can be trained end-to-end. Extensive experimental evaluation has been performed on both synthetic and real data including real photographs and computer tomography (CT) images. The results demonstrate that our model outperforms the previous models trained on unpaired images as well as the state-of-the-art methods based on paired training data when proper training pairs are unavailable.

Introduction

Image denoising is a classic low level vision problem but remains as a research hotspot because it is essential in various image processing and computer vision tasks. Noise corruption is usually inevitable when images are generated, which may heavily degrade the image quality. Image denoising aims at restoring the noise-free image from a noisy observation by reducing the latent noise. In many real situations, noise is generated by a very complicated process. For example, the noise in real photographs may be affected by the in-camera processing pipeline and environmental factors such as low illumination, radial distortion, over-exposure, *etc*. For medical CT images, their noise may be even more complicated, which is closely related to the radiation dose, resolution, slice thickness, patient size, organ type, *etc*. Therefore, it is very challenging to develop a general denoising method that is applicable to all noisy images.

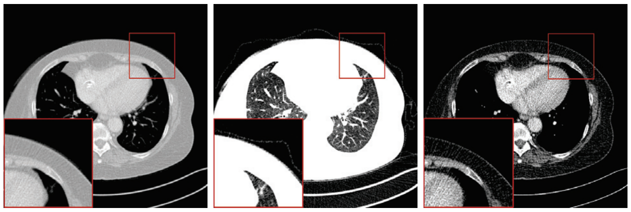

(a) [-1000,400] HU (b) [-1000,-600] HU (c) [-160,240] HU

Figure 1: An example low dose CT image normalized under different Hounsfield Unit (HU) ranges. The black area and highlighted area have very different noise distributions compared with other areas.

In the past few decades, many methods have been introduced in the literature trying to solve this problem. These methods can be roughly divided into two groups, image prior based models and discriminative learning based models. Image prior based methods such as BM3D (Dabov et al. 2007), NCSR (Dong et al. 2013) and WNNM (Zoran and Weiss 2011) are highly engineered approaches that are mostly based on self-similarity and have achieved impressive results for many years. Despite their good performance on some specific types of noise, they suffer from two main drawbacks. First, image priors used in these methods are mostly from human knowledge and experience, which are limited and not general enough to handle noise generated by complicated processes. Second, these methods involve complex optimization problems and may need to calculate the similarity between a large number of image patches, which is time-consuming.

To break the limitations of prior-based methods, several discriminative learning based methods (Chen and Pock 2017; Burger, Schuler, and Harmeling 2012; Schmidt and Roth 2014) have been proposed recently. They try to learn the latent noise implicitly from data and have achieved im-

[*]T. Jiang and J. Feng are the co-corresponding authors.

proved performance. Meanwhile, some Convolutional Neural Network (CNN) based methods, such as NLNet (Lefkimmiatis 2017) and DnCNN (Zhang et al. 2017), achieved even better results by training a deep neural network with paired clean and noisy data to remove additive white Gaussian noise (AWGN). In addition to AWGN, Guo et al. (2018) propose the CBDNet that uses a specially designed noise model for dealing with real photographs, achieving the state-of-the-art performance. However, the model does not extend to noise beyond that in real photographs. These discriminative learning based approaches learn noise from data directly, thus overcoming the limitations of prior-based methods. However, such discriminative models all require noisy and clean image pairs. Although it is possible to construct such image pairs for some noise such as AWGN, in most real scenarios including real photographs and low-dose CT images, such paired data are not always available or very hard to obtain. Therefore, it would be interesting to develop methods that are applicable when no paired training images are available.

If we could construct paired images from unpaired ones, then we would be able to perform image denoising similar to the existing methods based on paired data. There are a few methods adopting this idea such as GCBD (Chen et al. 2018), which trains a GAN model to learn the underlining noise and add it to clean data to construct image pairs, hence requiring specially selected image patches to model the noise distribution. This work assumes that the noise has zero-mean and extracts patches with similar internal content and weak background from the training images. A critical drawback of this method is that noise is implicitly assumed to be independent of image content, which does not hold in many cases (Foi et al. 2008; Liu, Tanaka, and Okutomi 2014), including, *e.g.* CT images (Fig. 1). Moreover, the noise generator learned on such patches often cannot generate proper noise for the foreground.

Hence, the above denoising methods either require paired training data or impose some limitations on noise distribution and image types. Is it possible to solve the blind image denoising problem where paired training data is not provided without any assumptions on data? To answer this question, we combine a cGAN with a image sharpening technique to generate image noise with any distribution that could be dependent of the image content. The noise generated by our noise generation component is further added to clean images to simulate noisy images. This results in clean and pseudo-noisy image pairs. Such pairs are used to train a denoising network similar to the existing methods based on paired images. The noise generation component and denoising network are integrated together so that they can be trained end-to-end. The whole framework is trained on unpaired clean and noisy images. More details will be given in the Proposed Method section. We evaluate the proposed method on both synthetic and real world data including real photographs and CT images. The results demonstrate that our method outperforms the existing methods trained on unpaired images as well as the state-of-the-art methods based on paired training data when proper training pairs are unavailable (*e.g.*, when denosing low-dose CT images).

The major contributions of our method include but not limited to: (1) We proposed a general end-to-end framework for image denoising without paired supervision. (2) Since our model does not make any special assumption on noise distribution and data types, it performs well on complex data such as medical images with content-dependent noise. The extensive experimental results demonstrate the superior performance of our model compared with the state-of-the-art image denoising methods. (3) We introduced an image sharpening technique for the GAN model to better capture image textural information. This technique has potential applications in problems beyond image denoising such as single image super-resolution and image style transfer.

Related Work

Prior-based image denoising Before discriminative learning models were introduced, various methods were proposed to model image priors such as models based on filters (Dabov et al. 2007), models based on sparse coding (Mairal et al. 2009; Elad and Aharon 2006; Dong et al. 2013), effective prior models (Zoran and Weiss 2011), low rank models (Gu et al. 2014) and models based on Markov Random Fields (MRFs) (Lan et al. 2006). Particularly, the self-similarity driven techniques among them, such as BM3D (Dabov et al. 2007), NCSR (Dong et al. 2013) and WNNM (Gu et al. 2014), have achieved impressive performance for many years. To denoise, the above methods treat each image independently without the requirement of a large training dataset. Therefore, they do not take into account the shared noise and content information between similar images. Furthermore, image priors based on human knowledge are often not general enough to model complicated noise. For methods such as BM3D, the self-similarity calculation between different similar patches also incur computation inefficiency.

Deep neural networks for image denoising Recent methods based on deep neural networks have outperformed traditional prior based methods. These methods follow a data-driven paradigm instead of relying only on analytical operations. Jain and Seung (2009) used a CNN for image denoising and observed that CNNs are more capable in representation learning than MRFs. Burger, Schuler, and Harmeling (2012) used a multi-layer perceptron (MLP) to denoise images which as we know is the first discriminative learning model to achieve comparative results as BM3D. Xie, Xu, and Chen (2012) used stacked sparse auto-encoders to deal with Gaussian noise. Chen and Pock (2017) proposed a dynamic trainable nonlinear reaction diffusion (TNRD) model with time-dependent parameters. Apart from these relatively early CNN methods, Zhang et al. (2017) proposed a deep CNN denoising model DnCNN using residual learning and batch normalization strategies, which is very powerful in handling Gaussian noise. Guo et al. (2018) proposed another convolutional blind denoising model (CBDNet) for real photographs that achieved the state-of-the-art performance. In this work, the authors considered both the signal-dependency and in-camera processing pipeline to better model noise on real photographs. However, the good performance achieved by the above methods are all based on the

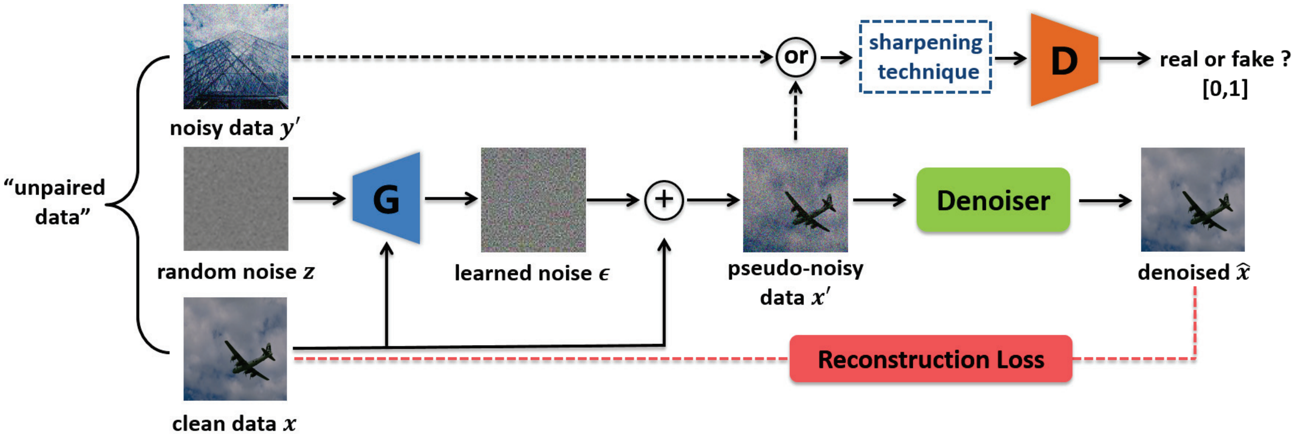


Figure 2: An overview of the proposed UIDNet framework. Given unpaired images, we first use a cGAN to learn the noise distribution from the noisy data. The noise obtained from the generator (denoted as G) is added to the clean data to form pseudo-noisy images. The discriminator (denoted as D) tries to distinguish the generated pseudo-noisy images from the real noisy images. This gives us pairs of clean and pseudo-noisy images. Such pairs are further used to train the denoiser. The cGAN and denoiser are integrated together so that they can be trained end-to-end.

premise that a large paired training dataset is available. But the premise does not hold in many real situations and thus these methods would fail. When this happens, one would have to seek help from methods based on unpaired data.

Image denoising without paired data There are few methods proposed to deal with image denoising in the absence of paired training data. Lehtinen et al. (2018) proposed a Noise2Noise model that learns image restoration without clean data under certain suitable and common circumstances based on statistical techniques. The limitation of this method is that the denoised image is not sharp enough compared with models trained on clean data. The other noise modeling technique proposed by Chen et al. (2018) in (GCBD) may be applied to images beyond real photographs. However, GCBD implicitly assumes that the noise has zero-mean and the noise from different patches of the images follows similar distributions.

Low-dose CT image denoising X-ray is one of the most widely utilized imaging modalities that has shown its great capabilities in disease diagnosis in modern hospitals and clinics (Brenner and Hall 2007). However, the potential cancer risk of X-ray radiation exposure to patients has raised public concerns (De González et al. 2009) due to the increasing use of medical CT scans. Therefore, low-dose CT has attracted widespread attention in medical imaging field and is preferred by more and more patients. Nevertheless, lowering the radiation dose also introduces certain noise and artifacts in reconstructed images that may compromise diagnostic performance. Hence, low-dose CT image reconstruction and denoising have become essential and many efforts have been made to deal with the involved technical issues, among which image post-processing methods have attracted considerable interest. Chen et al. (2017b) trained a deep CNN to transform low-dose CT images to normal-dose CT images patch by patch. Chen et al. (2017a) used a residual encoder-decoder network to perform low-dose CT denoising with impressive success. Yang et al. (2018) applied WGAN (Arjovsky, Chintala, and Bottou 2017) and perceptual loss to reduce noise and showed that GANs are superior than general CNN models in keeping image details. You et al. (2018) proposed a novel 3D noise reduction method on low-dose CT images. However, the above methods are based on paired low-dose and normal-dose training data, which are not always available. In this paper, we will apply our model to this problem and show the superior performance of our model over the existing applicable methods. The detailed results are given in the Experiments section.

Proposed Method

The goal of image denoising is to learn a function that maps the input noisy image to its noise-free version. Our workflow starts with an unpaired clean image set X and noisy image set Y', containing training samples $\{x_i\}_{i=1}^{N}$ where $x_i \in X$ and $\{y'_j\}_{j=1}^{M}$ where $y'_j \in Y'$. Different from most neural network based denoising methods due to the lack of paired training data, we first build paired images from given unpaired clean and noisy images, with which we can further train another denoising network similar to the existing methods based on paired images. Therefore, how to construct paired data is the crucial step of our model. To address this, we train a cGAN to learn the noise distribution in Y' and transform the clean data X to its corresponding pseudo-noisy version X' without making any assumption on the distribution of the noise. In this way, we construct the clean and pseudo-noisy image pairs $\{X, X'\}$ as desired. The noise learning network and the denoising network are integrated together so that they can be trained end-to-end. We denote the whole model as UIDNet (Unpaired Image Denosing Network). An overview of UIDNet is shown in Fig. 2. More details will be discussed in the following subsections.

The Noise Learning Network

The noise learning network is a Generative Adversarial Network (GAN). It learns the noise distribution in the noisy data Y' and transforms the clean data X to its corresponding pseudo-noisy version X'. It consists of a generative network G and a discriminative network D. The generator G is trained to generate samples that are closed to real data from a random noise sample and D is trained to distinguish whether a sample is generated by G or from real data. In the original GAN, D and G are trained to solve the following minimax optimization problem

$$\begin{aligned}\min_{G}\max_{D}\mathcal{L}_{GAN}(D,G) &= \mathbb{E}_{x\sim P_{data}(x)}\left[log(D(x))\right] \\ &+ \mathbb{E}_{z\sim P_{z}(z)}\left[log(1-D(G(z)))\right]\end{aligned} \tag{1}$$

where $\mathbb{E}(\cdot)$ denotes the expectation operator, and P_{data} and P_z are the distributions of real data and random noise. The plain GAN generates images just from a random noise sample z, which is not a proper option for us here because noise is often related to image content in most real world situations. Therefore, we have to revise the generator such that the generated noisy image X' is related to the clean data X. Mirza and Simon proposed a conditional version of GAN, called cGAN (Mirza and Osindero 2014), to deal with this kind of problem. In cGAN, the generator could generate samples based on some condition c, which could be any kind of auxiliary information, such as class labels or data from other modalities. In our problem, we treat a clean image x as the condition c and perform the conditioning by feeding it into the generator as an additional input apart from the random sample z.

Although the classic GAN model such as DCGAN (Radford, Metz, and Chintala 2015) has proved its capability in learning data distributions, it relies on minimizing the Jensen-Shannon (JS) divergence between the distributions of the generated and real data, which may suffer from vanished gradient on the generator G (Arjovsky, Chintala, and Bottou 2017) in certain circumstances where G may stop updating its parameters during training. This motivated the introduction of WGAN (Arjovsky, Chintala, and Bottou 2017) based on the Wasserstein distance, which has better performance in generating images and is easier to train. In our model, an improved version of WGAN, called WGAN-GP (Gulrajani et al. 2017), is applied to learn the noise distribution with the following objective function

$$\begin{aligned}\min_{G}\max_{D}\mathcal{L}_{WGAN}(D,G) &= \mathbb{E}_{y'\sim P_{Y'}}[D(y')] \\ &-\mathbb{E}_{z\sim P_z, x\sim P_X}\left[D(G(z,x))\right] \\ &+\lambda\mathbb{E}_{\hat{x}\sim P_{\hat{x}}}\left[(||\nabla_{\hat{x}}D(\hat{x})||_2-1)^2\right]\end{aligned} \tag{2}$$

where P_z , $P_{Y'}$ and P_X are the distributions of random noise, real noisy data Y' and real clean data X'. $P_{\hat{x}}$ is the distribution of $\hat{x}$ that is sampled uniformly along straight lines between pairs of generated and real samples (Gulrajani et al. 2017). WGAN-GP removes the log term in its loss function and adds a gradient penalty term for network regularization compared with the original GAN. Particularly, it drops the last sigmoid layer of the discriminator in the implementation. The detailed architecture of the WGAN-GP network is described below.

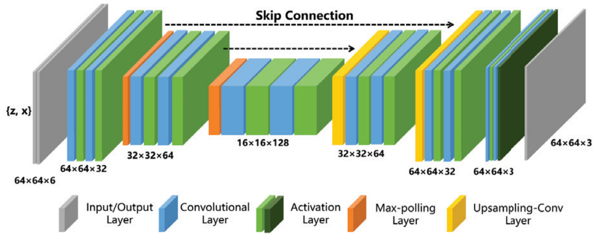


Figure 3: The generator architecture of the proposed UID-Net.

Generative Network Architecture The generator is a U-Net (Ronneberger, Fischer, and Brox 2015) like network that takes a clean image x and random noise sample z as two channels of the input. It is widely known that the training of a GAN model tends to be unstable. Hence, it is difficult for the network to directly generate high quality pseudo-noisy images X' with noise similar to the given noisy data Y' without introducing artifacts. To ease the training process, we set the generator to generate only noise and then add it to the clean image X to obtain the X'. Here we implicitly assumed that all the noisy images have the same noise distribution. The generator architecture is illustrated in Fig. 3. It consists of an encoder network and a decoder network with skip connections between them. The encoder part just follows the general CNN architecture with repeated conv-pool units, *i.e.*, each unit consists of two 3×3 convolutional layers followed by a rectified linear unit (ReLU) (He et al. 2015) activation layer and a 2×2 max-pooling layer for downsampling. The number of feature channels in the first convolutional layer is 32 and it doubles at each max-pooling layer. The decoder part performs opposite processing with upsampling or deconvolutional layers. Here, we use an "up-conv" operation consisting of an upsampling layer followed by a 2×2 convolutional layer to double the feature map size and halve the number of feature channels. Then, with a skip connection, it is concatenated with the corresponding layers in the encoder part, and is followed by two 3×3 convolutional layers and a ReLU activation layer. At the end, a 1×1 convolutional layer is used to output the target image with the desired number of channels. This is also referred to as a "fully convolutional network" (Long, Shelhamer, and Darrell 2015) that could handle images with different input sizes.

Discriminative Network Architecture The discriminator takes a real noisy image from Y' or generated pseudo-noisy image from X' as input and returns the probability that the image is sampled from real noisy data. The architecture of the discriminator is illustrated in Fig. 4 which is similar to DCGAN (Radford, Metz, and Chintala 2015). The difference is that we have removed the batch normalization layers and the last sigmoid layer as WGAN-GP (Gulrajani et al. 2017). In the training process, we use 64×64 image patches to train the GAN network. The images fed to the

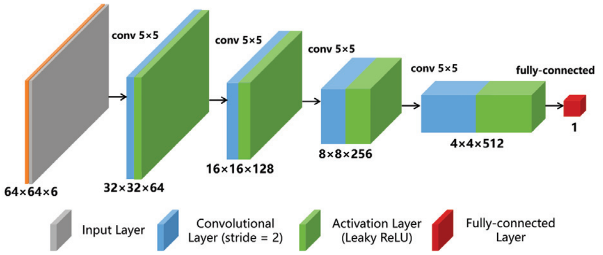


Figure 4: The discriminator architecture of the proposed UIDNet.

discriminator are X' and Y' that may have different image contents. To make the discriminator focus more on noise, we apply a sharpening technique by subtracting a image from its local-mean filtered version. The sharpened images and original images serve as two group of channels of the input for the discriminator. This sharpening technique is critical to the success of model in performance. The kernel size for calculating local mean is a hyper-parameter and is set to 3 × 3 in our experiments.

The Denoising Network

The cGAN in above subsection helps us to construct the paired clean and pseudo-noisy data $\{X, X'\}$. Now we perform image denoising similar to the existing paired methods. As mentioned before, there are many neural network based models (Lefkimmiatis 2017; Zhang et al. 2017) proposed for this with greater capabilities than prior based models. Here, we adopt a network architecture similar to DnCNN (Zhang et al. 2017). Since the batch normalization (Ioffe and Szegedy 2015) and residual learning (He et al. 2016) strategies are known to help improve the performance in image denoising. We incorporate them here too. Our network takes a noisy image x' as the input and outputs the residual noise ϵ that it has learned. The denoised clean image $\hat{x}$ is then obtained by subtracting the noise ϵ from the input noisy image x'. The involved objective function is

$$\mathcal{L}_{denoiser} = \frac{1}{2N}\sum_{i=1}^{N} ||(\hat{x_i} - x_i)||_F^2 \tag{3}$$

$$\hat{x_i} = x_i' - \epsilon_i \tag{4}$$

$$\epsilon_i = f_\Theta(x_i') \tag{5}$$

where N is the size of the training data, Θ denotes the network parameter we need to learn and the subscript i means the i-th sample of the data and F denotes Frobenius norm. The detailed network architecture is illustrated in Fig. 5. It consists of 16 repeated "conv-bn-relu" units, each of which contains a 3 × 3 convolutional layer followed by a batch normalization layer and a ReLU (He et al. 2015) activation layer except the first, similar to DnCNN (Zhang et al. 2017). At the end, an additional 3 × 3 convolutional layer is used to output the target noise image with the desired number of channels. All the convolutional layers adopt the zero padding strategy to keep the feature map size consistent. Again, it forms a "fully convolutional network" (Long, Shelhamer, and Darrell 2015) that allows us to test our model on images of different sizes after training. Finally, the noise learning and denoising modules are integrated together and trained end-to-end jointly.

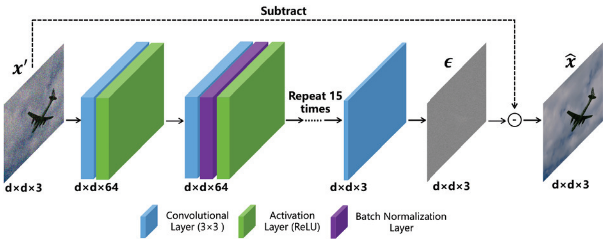


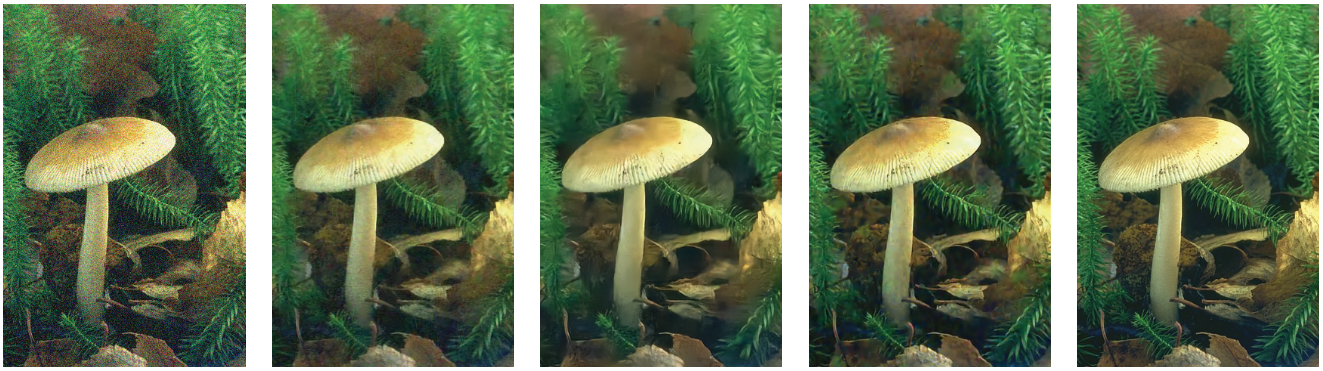
Figure 5: The architecture of the denoising network.

Experiments

In this section, we evaluate our UIDNet's performance on both synthetic and real world data including real photographs and medical CT images. Our experiments are divided into three parts and UIDNet is compared with several representative published methods in each part. (1) We evaluate our model's performance on synthetic data with independent Gaussian noise to confirm that it is able to handle such simple noise with unpaired data. (2) We evaluate our model's performance on real photographs to show that it is capable of dealing with real world noise. (3) Our model is applied to denoise low-dose CT image to demonstrate our model's capability in handling even more complicated noise. In the following, we denote our method with the sharpening technique as UIDNet and the version without the technique as UIDNet-NS.

Experimental Setting

Training and Test Data To train and test the model, we crop images into 64 × 64 patches in all experiments. (1) For denoising synthethic Gaussian noise, we use the 400 images of size 180 × 180 in (Chen and Pock 2017) for training. The noisy images are obtained by adding Gaussian noise to these images. We consider three representative noise levels as DnCNN (Zhang et al. 2017), *i.e.*, $\sigma = 15$, 25 and 50. We crop 25, 600 image patches to train our model. After training, we test our model on a popular test data BSD68 including 68 natural images from the Berkeley segmentation dataset (Roth and Black 2009). Apart from these gray images, we also train our model on color images from the BSD500 dataset (Arbelaez et al. 2011). We use the color version of BSD68 (denoted as CBSD68) as the test images and the remaining 432 color images as the training images. In total, 63, 000 image patches are extracted. (2) For image denoising on real photographs, we choose the benchmark Smartphone Image Denoising Dataset (SIDD) (Abdelhamed, Lin, and Brown 2018), which consists of 30, 000 noisy images and correspoding high-quality ground truth images. The images are from 10 different scenes under different lighting conditions using five representative smartphone cameras (Apple iPhone 7, Google Pixel, Samsung



(a) Noisy / 20.44 dB (b) BM3D / 28.10 dB (c) DnCNN / 27.46 dB (d) UIDNet / 28.23 dB (e) Ground truth

Figure 6: The denoising results in PSNR on an image from CBSD68 with Gaussian noise $\sigma = 25$.

Table 1: The average PSNR(dB) results of the compared methods on BSD68 in denoising with Gaussian synthetic noise.

Method / Noise	Paired / Supervision			Unpaired / Without paired supervision				
	MLP	TNRD	DnCNN	BM3D	WNNM	EPLL	UIDNet-NS	UIDNet
$\sigma = 15$	-	31.42	31.61	31.07	31.37	31.21	30.41	31.30
$\sigma = 25$	28.96	28.92	29.16	28.57	28.83	28.68	28.05	28.98
$\sigma = 50$	26.03	25.97	26.23	25.62	25.87	25.67	25.29	26.04

Galaxy S6 Edge, Motorola Nexus 6 and LG G4). The ground truth of this dataset is estimated with a sophisticated processing pipeline including defective pixel correction, intensity alignment, dense local spatial alignment and robust mean image estimation. The SIDD dataset has higher quality and is better than the RENOIR dataset (Anaya and Barbu 2018) and the Darmstadt Noise Dataset (DND) (Plotz and Roth 2017). Note that although the DND has been used in previous work (Chen et al. 2018; Guo et al. 2018), it is not suitable for benchmarking our model since it does not provide training data. For the purpose of benchmarking, the authors of (Abdelhamed, Lin, and Brown 2018) pick 200 images with one for each scene instance from the SIDD, where 40 representative images are used as the test data and the remaining 160 noisy images and their ground truth images are made available for training. For the efficiency of evaluation, 32 randomly selected non-overlapping image patches of size 256×256 from each of the 40 test images are provided, forming a total of 1280 test image patches. We adopt the same training and test dataset split strategy and crop $520,965$ image patches from the 160 training images to train UIDNet. (3) For low-dose CT image denoising, a real clinical dataset from *"the 2016 NIH-AAPM-Mayo Clinic Low Dose CT Grand Challenge"* authorized by Mayo Clinic (AAPM 2016) is utilized to train and evaluate our model. This dataset consists of 5936 slices of 1mm thickness and 2378 slices of 3mm thickness normal-dose (full dose) and low-dose (quarter dose) abdominal CT images with size of 512×512 from 10 anonymous patients. We denote the 3mm thickness CT images as set A and 1mm thickness CT images as set B. We denote low-dose CT images as Noisy and normal-dose CT images as Clean. Therefore, we have four combinations: CleanA, NoisyA, CleanB and NoisyB. For image preprocessing, we randomly extract overlapping patches and exclude completely black patches corresponding to air area in CT image. Finally, CleanA (or NoisyA) are splitted into training and testing sets with $197,214$ and 5234 patches, respectively. CleanB (or NoisyB) are splitted into training and testing sets with $198,796$ and 4768 patches, respectively. The values of Hounsfield Unit (HU) on the CT images are normalized to [0,1] according to the abdominal window width of [-160,240] HU as Yang et al. (2018).

Implementation Details Before each epoch of the training process, all the clean and noisy image patches are shuffled. In each mini-batch, the clean and noisy images fed to the network are uncoupled. During the training phase, the batch size is set to 64. In the noise learning model, the λ in Eqn. 2 is set to 10 and local mean kernel size is set to 3×3. For the denoising network, we roughly follow the parameter settings in DnCNN (Zhang et al. 2017). We use the Adam (Kingma and Ba 2014) optimization algorithm with $\beta_1 = 0.5$ and initial learning rate 1.0×10^{-4} to train UIDNet. Depending on the training dataset size, we train the model for 100, 20 and 50 epochs on the synthetic data, real photographs and low-dose CT images, respectively. It takes seven to nine hours to train our model on a single Nvidia GeForce GTX 1080 Ti GPU.

Compared Methods In the experiment on synthetic data, we compare our model with BM3D (Dabov et al. 2007), WNNM (Gu et al. 2014), EPLL (Zoran and Weiss 2011), MLP (Burger, Schuler, and Harmeling 2012), TNRD (Chen and Pock 2017), and DnCNN (Zhang et al. 2017). For the real photographs, we also compare with the state-of-the-art model CBDNet (Guo et al. 2018). For low-dose CT images,

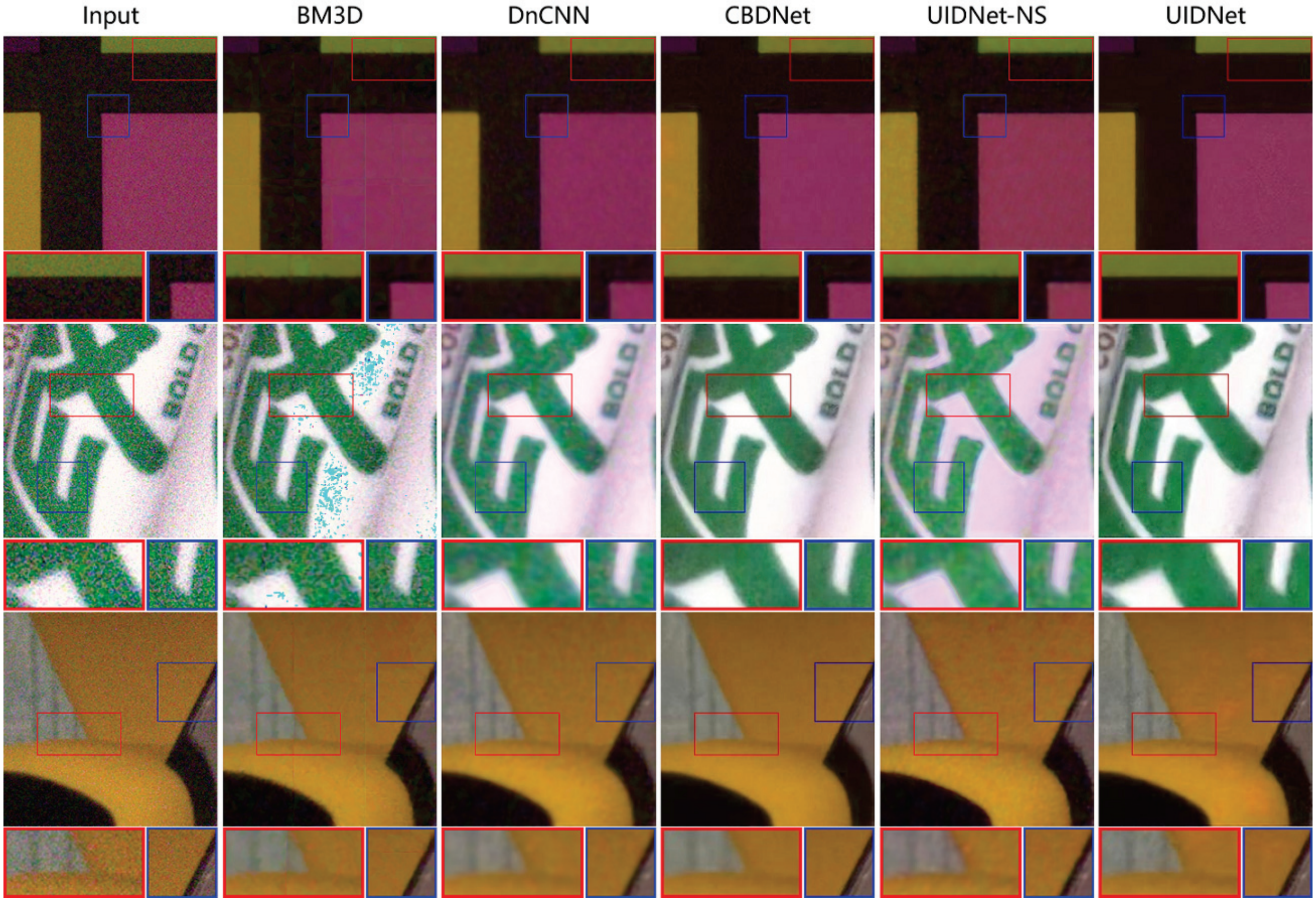


Figure 7: The denoising results of UIDNet on some example images from the dataset SIDD. Zoom in for a better view.

we compare our methods with several recently published models WGAN-VGG (Yang et al. 2018), SMGAN (You et al. 2018) and RED-CNN (Chen et al. 2017a). Unfortunately, we are unable to compare with GCBD (Chen et al. 2018) due to the unavailability of its source code. We use the popular PSNR (peak signal-to-noise ratio) and SSIM (structural similarity index) as quantitative measures of denoising performance.

Denoising on Synthetic Images

We evaluate the model's denoising performance on the BSD68 (Roth and Black 2009) dataset with Gaussian noise of zero mean and three representative standard deviations $\sigma = 15$, 25 and 50. The quantitative results are shown in Table 1. The results of the compared methods are taken from Zhang et al. (2017). In this task, although UIDNet was trained on unpaired images, it achieved a similar performance as MLP, TNRD and DnCNN that were trained on paired images. The effectiveness of the sharpening technique is clearly demonstrated by the gap between UIDNet-NS and UIDNet. One of the reasons for the overall small difference between the compared methods is that Gaussian noise is relatively easy to handle. We also train our model on color images. Fig. 6 shows an example of the denoising results on some CBSD68 image with Gaussian noise $\sigma = 25$, from which we can see that DnCNN oversmoothed the image and our model was able to keep more details.

Denoising on Real Photographs

Here, we evaluate our model on real photographs from the SIDD (Abdelhamed, Lin, and Brown 2018). We evaluate our model and the compared methods in the sRGB space for general image denoising. The quantitative denoising results are shown in Table 2 (mostly taken from (Abdelhamed, Lin, and Brown 2018)). Clearly, our method significantly outperforms traditional prior-based methods and the methods trained with paired data assuming Gaussian noise such as DnCNN. Our method even surpasses a state-of-the-art denoising method for real photographs, CBDNet, in terms of the SSIM index, although CBDNet was specially designed to denoise photographs that takes into account the in-camera image processing pipeline. We also noticed that the results of CBDNet are partially based on real noisy images and associated nearly noise-free images calculated by existing approaches. Such data provide extra information for the model. The large SSIM difference between UIDNet-NS and UIDNet again indicates that the sharpening technique is very effective for the network to capture structural information of

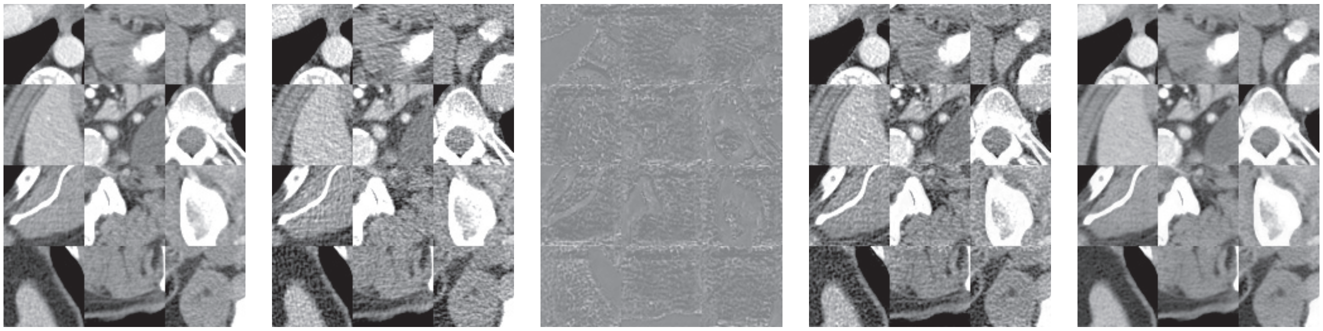

(a) Real NDCT (b) Real LDCT (c) Learned noise (d) Pseudo LDCT (e) Denoised output

Figure 8: Results of UIDNet on example CT patches. (a) Real normal-dose. (b) Real low-dose. (c) The learned noise. (d) The generated low-dose. (e) The denoised output from the real low-dose CT patches. The display window is [-160, 240] HU.

Table 2: The average PSNR(dB) and SSIM results of all compared methods on the SIDD in denoising real photographs.

Method		PSNR	SSIM
Paired / Supervision	MLP	24.71	0.641
	TNRD	24.73	0.643
	DnCNN	28.46	0.784
	CBDNet	33.28	0.868
Unpaired / Without paired supervision	BM3D	25.65	0.685
	KSVD-G	27.19	0.771
	NLM	26.75	0.699
	KSVD	26.88	0.842
	KSVD-DCT	27.51	0.780
	LPG-PCA	24.49	0.681
	FoE	25.58	0.792
	WNNM	25.78	0.809
	GLIDE	24.71	0.774
	EPLL	27.11	0.870
	UIDNet-NS	31.34	0.856
	UIDNet	32.48	0.897

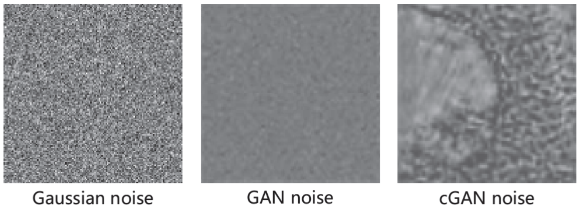


Figure 9: Comparison of Gaussian noise and the noises learned from an example CT patch by GAN and cGAN.

the images. Fig. 7 shows the denoising performance of UIDNet and some other popular denoising methods. Compared with BM3D and DnCNN, UIDNet was able to keep more sharp edges.

Low-dose CT Image Denoising

Finally, we evaluate our model on CT images from *"the 2016 NIH-AAPM-Mayo Clinic Low Dose CT Grand Challenge"* authorized by Mayo Clinic (AAPM 2016). Here, the normal-dose CT (NDCT) corresponds to clean images and low-dose CT (LDCT) corresponds to noisy images. Both images with z-spacing 3mm (dataset A) and z-spacing 1mm (dataset B) are considered. The quantitative results in PSNR and SSIM on dataset A are shown in Table 3 (the two columns under NoisyA). Note that, except UIDNet and BM3D, all other methods compared here require paired training data. Our model outperformed all other methods except RED-CNN in terms of PSNR, which suggests that our model does well on complex noise. Again, UIDNet consistently performed better than UIDNet-NS. Fig. 8 shows some examples where UIDNet was able to generate noise that is content dependent. Interestingly, because the model was not trained with paired data, it is possible for the denoised images Fig. 8(e) to look even smoother than the corresponding clean images. Fig. 8(a), Fig. 8(c) and Fig. 9 show that the cGAN could generate noise depending on the clean image content. Compared with Gaussian noise or noise generated by unconditioned GAN, such image dependent noise fits real data better. We have also done the ablation study on different realizations of initial random sample z. We tried the binomial distribution with probability 0.5 and observed a similar noise distribution. Because the noise generation is through a very complicated network and the noise distribution is mainly determined by the GAN, we think it makes sense that the initial z distribution has little influence on the resulted noise.

As mentioned in previous sections, most of the state-of-the-art denoising methods either require paired training data or impose some limitations on noise distribution and image types. However, in many real world situations such as medical image analysis, paired clean and noisy data could be very difficult to obtain. For example, in a normal workflow, patients would not take normal-dose CT images and low-dose ones at the same time, althrough it is possible to obtain NDCT and LDCT images on a limited set of patients and CT machines under proper ethics agreements. Mod-

Table 3: The quantitative PSNR and SSIM results of different methods on LDCT image denoising. The initial PSNR and SSIM of the LDCT images are 22.461 and 0.647 on NoisyA and are 17.702 and 0.568 on NoisyB, respectively.

Method		Training		Test			
		Clean	Noisy	NoisyA		NoisyB	
				PSNR	SSIM	PSNR	SSIM
Paired	WGAN-VGG	CleanA	NoisyA	25.300	0.722	20.236	0.462
	SMGAN	CleanA	NoisyA	25.507	0.732	20.354	0.653
	RED-CNN	CleanA	NoisyA	**27.243**	0.743	21.723	0.673
Unpaired	BM3D	-	-	26.325	0.728	21.439	0.661
	UIDNet-NS	CleanA	NoisyA	26.475	0.738	21.198	0.662
	UIDNet	CleanA	NoisyA	26.694	**0.746**	21.247	0.668
	UIDNet	CleanA	NoisyB	-	-	22.315	0.682
	UIDNet	CleanA	NoisyB & NoisyA	-	-	**22.578**	**0.686**

els trained on such paired data can be further applied to other patients on other machines with different noise levels and distributions. To simulate this situation, we hide the NDCT images in dataset B (*i.e.,* CleanB) from the paired methods and train them with the NDCT and LDCT images in dataset A (*i.e.,* CleanA and NoisyA). On the other hand, we train the unpaired methods with CleanA and the LDCT images in dataset B (*i.e.,* NoisyB). All methods are tested on NoisyB using CleanB as the ground truth. The test results are shown in Table 3 (the last two columns under NoisyB). UIDNet (with PSNR 22.315 and SSIM 0.682) did remarkably well on NoisyB, outperforming the state-of-the-art paired methods including RED-CNN. The effectiveness of the sharpening technique is further demonstrated in the better performance of UIDNet over UIDNet-NS when trained on dataset A. The performance of UIDNet when trained on both NoisyA and NoisyB is also provided in the table for reference.

Conclusion

In this paper, we proposed an end-to-end blind image denoising framework consisting of a noise learning network based on cGANs and a denoising network based on CNNs. Compared with the existing methods, the model imposes the weakest assumptions on noise distribution and data types. The most critical part of the model is to generate image content dependent noises. To make it possible, we used a cGAN with a U-Net type generator. A sharpening technique was introduced to further improve the performance of the model. With properly generated noise conditioned on clean image, we constructed clean and pseudo-noisy image pairs to train a denoising network similar to the previous methods based on paired images. All the components were integrated together so that they could be trained end-to-end. Extensive evaluation was performed on both synthetic and real world data including real photographs and CT images. The results demonstrate that for synthetic Gaussian noise, our model's performance is close to the previous methods based on paired data. On real photographs, our model significantly outperformed the previous prior-based methods such as BM3D and representative discriminative learning based methods such as DnCNN, and it performed comparably to a state-of-the-art method for denoising real photographs, CBDNet. On low-dose CT images with more complex noise, our model also showed its capability in generating content dependent noise and achieved better denoising performance than the state-of-the-art methods based on paired training data when proper paired training images are unavailable. Moreover, we believe that the proposed unpaired UIDNet framework and sharpening technique have potential applications in solving other related problems such as single image super-resolution and image style transfer.

Acknowledgements: This work has been supported in part by the National Natural Science Foundation of China grant 61772197, the National Key Research and Development Program of China grant 2018YFC0910404 and Beijing Natural Science Foundation grant 4192044.

References

AAPM. 2016. Low dose ct grand challenge. https://www.aapm.org/GrandChallenge/LowDoseCT/. 2016.

Abdelhamed, A.; Lin, S.; and Brown, M. S. 2018. A high-quality denoising dataset for smartphone cameras. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 1692–1700.

Anaya, J., and Barbu, A. 2018. Renoir–a dataset for real low-light image noise reduction. *Journal of Visual Communication and Image Representation* 51:144–154.

Arbelaez, P.; Maire, M.; Fowlkes, C.; and Malik, J. 2011. Contour detection and hierarchical image segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* 33(5):898–916.

Arjovsky, M.; Chintala, S.; and Bottou, L. 2017. Wasserstein generative adversarial networks. In *International Conference on Machine Learning*, 214–223.

Brenner, D. J., and Hall, E. J. 2007. Computed tomography–an increasing source of radiation exposure. *New England Journal of Medicine* 357(22):2277–2284.

Burger, H. C.; Schuler, C. J.; and Harmeling, S. 2012. Image denoising: Can plain neural networks compete with bm3d? In *2012 IEEE conference on computer vision and pattern recognition*, 2392–2399. IEEE.

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, H.; Zhang, Y.; Kalra, M. K.; Lin, F.; Chen, Y.; Liao, P.; Zhou, J.; and Wang, G. 2017a. Low-dose ct with a residual encoder-decoder convolutional neural network. *IEEE transactions on medical imaging* 36(12):2524–2535.

Chen, H.; Zhang, Y.; Zhang, W.; Liao, P.; Li, K.; Zhou, J.; and Wang, G. 2017b. Low-dose ct denoising with convolutional neural network. In *2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017)*, 143–146. IEEE.

Chen, J.; Chen, J.; Chao, H.; and Yang, M. 2018. Image blind denoising with generative adversarial network based noise modeling. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 3155–3164.

Dabov, K.; Foi, A.; Katkovnik, V.; and Egiazarian, K. 2007. Image denoising by sparse 3-d transform-domain collaborative filtering. *IEEE Transactions on Image Processing* 16:2080–2095.

De González, A. B.; Mahesh, M.; Kim, K.-P.; Bhargavan, M.; Lewis, R.; Mettler, F.; and Land, C. 2009. Projected cancer risks from computed tomographic scans performed in the united states in 2007. *Archives of internal medicine* 169(22):2071–2077.

Dong, W.; Zhang, L.; Shi, G.; and Li, X. 2013. Nonlocally centralized sparse representation for image restoration. *IEEE Transactions on Image Processing* 22(4):1620–1630.

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 *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2862–2869.

Gulrajani, I.; Ahmed, F.; Arjovsky, M.; Dumoulin, V.; and Courville, A. C. 2017. Improved training of wasserstein gans. In *Advances in Neural Information Processing Systems*, 5767–5777.

Guo, S.; Yan, Z.; Zhang, K.; Zuo, W.; and Zhang, L. 2018. Toward convolutional blind denoising of real photographs. *arXiv preprint arXiv:1807.04686*.

He, K.; Zhang, X.; Ren, S.; and Sun, J. 2015. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In *Proceedings of the IEEE international conference on computer vision*, 1026–1034.

He, K.; Zhang, X.; Ren, S.; and Sun, J. 2016. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 770–778.

Ioffe, S., and Szegedy, C. 2015. Batch normalization: Accelerating deep network training by reducing internal covariate shift. *arXiv preprint arXiv:1502.03167*.

Jain, V., and Seung, S. 2009. Natural image denoising with convolutional networks. In *Advances in neural information processing systems*, 769–776.

Kingma, D. P., and Ba, J. 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

Lan, X.; Roth, S.; Huttenlocher, D.; and Black, M. J. 2006. Efficient belief propagation with learned higher-order markov random fields. In *European conference on computer vision*, 269–282. Springer.

Lefkimmiatis, S. 2017. Non-local color image denoising with convolutional neural networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 3587–3596.

Lehtinen, J.; Munkberg, J.; Hasselgren, J.; Laine, S.; Karras, T.; Aittala, M.; and Aila, T. 2018. Noise2noise: Learning image restoration without clean data. *arXiv preprint arXiv:1803.04189*.

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.

Long, J.; Shelhamer, E.; and Darrell, T. 2015. Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 3431–3440.

Mairal, J.; Bach, F.; Ponce, J.; Sapiro, G.; and Zisserman, A. 2009. Non-local sparse models for image restoration. In *2009 IEEE 12th International Conference on Computer Vision (ICCV)*, 2272–2279. IEEE.

Mirza, M., and Osindero, S. 2014. Conditional generative adversarial nets. *arXiv preprint arXiv:1411.1784*.

Plotz, T., and Roth, S. 2017. Benchmarking denoising algorithms with real photographs. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 1586–1595.

Radford, A.; Metz, L.; and Chintala, S. 2015. Unsupervised representation learning with deep convolutional generative adversarial networks. *arXiv preprint arXiv:1511.06434*.

Ronneberger, O.; Fischer, P.; and Brox, T. 2015. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, 234–241. Springer.

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

Schmidt, U., and Roth, S. 2014. Shrinkage fields for effective image restoration. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2774–2781.

Xie, J.; Xu, L.; and Chen, E. 2012. Image denoising and inpainting with deep neural networks. In *Advances in neural information processing systems*, 341–349.

Yang, Q.; Yan, P.; Zhang, Y.; Yu, H.; Shi, Y.; Mou, X.; Kalra, M. K.; Zhang, Y.; Sun, L.; and Wang, G. 2018. Low-dose ct image denoising using a generative adversarial network with wasserstein distance and perceptual loss. *IEEE transactions on medical imaging* 37(6):1348–1357.

You, C.; Yang, Q.; Gjesteby, L.; Li, G.; Ju, S.; Zhang, Z.; Zhao, Z.; Zhang, Y.; Cong, W.; Wang, G.; et al. 2018. Structurally-sensitive multi-scale deep neural network for low-dose ct denoising. *IEEE Access* 6:41839–41855.

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

Zoran, D., and Weiss, Y. 2011. From learning models of natural image patches to whole image restoration. In *2011 International Conference on Computer Vision*, 479–486. IEEE.