NTGAN: Learning Blind Image Denoising without Clean Reference

Rui Zhao
rick10.zhao@connect.polyu.hk

Daniel P.K. Lun
enpklun@polyu.edu.hk

Kin-Man Lam
enkmlam@polyu.edu.hk

Department of Electronic and
Information Engineering,
The Hong Kong Polytechnic University,
Hong Kong, China

Abstract

Recent studies on learning-based image denoising have achieved promising performance on various noise reduction tasks. Most of these deep denoisers are trained either under the supervision of clean references, or unsupervised on synthetic noise. The assumption with the synthetic noise leads to poor generalization when facing real photographs. To address this issue, we propose a novel deep unsupervised image-denoising method by regarding the noise reduction task as a special case of the noise transference task. Learning noise transference enables the network to acquire the denoising ability by only observing the corrupted samples. The results on real-world denoising benchmarks demonstrate that our proposed method achieves state-of-the-art performance on removing realistic noises, making it a potential solution to practical noise reduction problems.

1 Introduction

Convolutional neural networks (CNNs) have been widely studied over the past decade in various computer vision applications, including image classification [13], object detection [29], and image restoration [35]. In spite of high effectiveness, CNN models always require a large-scale dataset with reliable reference for training. However, in terms of low-level vision tasks, such as image denoising and image super-resolution, the reliable reference data is generally hard to access, which makes these tasks more challenging and difficult. Current CNN-based denoising methods usually approximate the realistic noise as the additive white Gaussian noise (AWGN), and synthesize the pairs of the corrupted observation $\mathbf{y}_i$ and the clean reference $\mathbf{x}_i$ for training. With a large amount of these synthetic pairs, a deep regression model, i.e. a convolutional neural network, can be trained by minimizing the empirical risk, as follows:

$$\Theta^* = \arg\min_{\Theta} \sum_i \mathcal{L}(f(\mathbf{y}_i; \Theta), \mathbf{x}_i), \tag{1}$$

where f denotes a mapping model from $\mathbf{y}_i$ to $\mathbf{x}_i$ with its trainable parameters Θ under a loss function $\mathcal{L}$. However, in some practical applications, e.g., medical imaging and hyperspectral remote sensing, the acquisition of noise-free data is expensive, or even impractical.

In addition, most state-of-the-art image denoising methods require *a priori* knowledge of the noise level when performing inference. Methods, based on image prior modeling [9, 18], usually involve some optimization processes, and the hyperparameters for solving these optimization problems are highly related to the noise level. On the other hand, methods, based on plain discriminative learning, attempt to either learn a denoising function for a specific noise level, e.g., DnCNN [55], or take the noise level as an input variable when training the models, e.g., FFDNet [56] and CBDNet [10]. Therefore, the noise level plays an important role in practical noise reduction tasks. However, current deep unsupervised image denoisers lack the utilization of the information provided by the noise level, which results in poor generalization and flexibility when solving practical denoising problems.

In this paper, we propose a novel noise transference generative adversarial network, namely NTGAN, to tackle the problems of limited reference data and poor generalization in denoising real-world photographs. Specifically, NTGAN attempts to learn the noise transference between the paired noisy observations. We follow the idea of conditional generative adversarial network [22] and establish a controllable generator for different noise levels.

In summary, the main contributions of this work can be concluded as follows:

- We treat noise reduction as a special case of noise transference, and establish a conditional regression model with the guidance from the target noise level, which enables the proposed denoiser to be trained by observing corrupted samples only.

- We present a strategy for generating paired training samples from real photographs with the pseudo supervision from the target noise level maps, which allows the network to obtain better generalization ability for removing realistic noises.

2 Related work

2.1 Blind image denoising

Blind image denoising is a challenging and complicated task, because the realistic noise is spatially variant and signal-dependent, which makes the noise distribution hard to estimate. Methods, based on image prior modeling, usually approximate the realistic noise as AWGN and estimate the noise level via principal component analysis [5, 21, 28] and wavelet transform [27] for effective noise removal. Recently, more and more attention has been paid to the extension of AWGN. Numerous studies have been proposed to investigate the noise correlations between image channels [51] and image patches [53]. On the other hand, methods, based on plain discriminative learning, are generally trained with the synthetic pairs based on AWGN [7, 55]. A great breakthrough comes from a two-stage denoising system, i.e. CBDNet [10], which aims to estimate the pixel-wise noise level via a sub-network, followed by a non-blind deep denoiser for removing the noise. VDNet [54] further enhances the noise model in CBDNet, and achieves the state-of-the-art performance. RIDNet [3], from another aspect, proposed that the "residual-on-residual" architecture can perform blind image denoising in a one-stage manner.

2.2 Unsupervised image denoising

In spite of their great power of removing noise, the learning-based methods usually require a large amount of noisy and noise-free image pairs for training. However, the noise-free

references are generally difficult to obtain in practice, which leads to the challenge of unsupervised image denoising. Noise2Noise (N2N) [17] proposed that the models, learnt from the paired corrupted observations, can be applied to effectively removing the noise. In addition, Noise2Self (N2S) [4] and Noise2Void (N2V) [14] adopted the self-supervision strategy to train a regression model for blind noise removal. From another point of view, the work of deep image prior (DIP) [30] showed that the network architecture itself contains *a priori* knowledge, which can be used for image restoration. However, the above-mentioned works mainly consider the synthetic noise, which results in a dramatic degradation of their performance when facing real photographs. Therefore, in this paper, we move one step further to establishing an unsupervised denoising framework for the realistic noise.

3 The proposed noise transference GAN

In this section, we present the proposed NTGAN for blind unsupervised image denoising. Firstly, we analyze the relationship between noise transference and noise reduction. Secondly, we introduce the proposed pseudo noise level map for unsupervised learning. Finally, we present the learning strategy for training the proposed NTGAN.

3.1 Noise transference

Conventional deep denoisers [35, 33] attempt to learn a denoising function for a specific noise level. Therefore, the resultant model parameters Θ become a function of the noise level σ as $\hat{\mathbf{x}} = f(\mathbf{y}; \Theta(\sigma))$, where $\hat{\mathbf{x}}$ is the estimated noise-free image, and $\mathbf{y}$ is the noisy observation with the noise level σ. As claimed in FFDNet [36], the model flexibility can be enhanced by introducing the noise level σ as an input variable into the model. Thus, the denoising function becomes $\hat{\mathbf{x}} = f(\mathbf{y}, \sigma; \Theta)$. During training, the following objective function is minimized:

$$\Theta^* = \arg\min_{\Theta} \sum_i \mathcal{L}(f(\mathbf{y}_i, \sigma_i; \Theta), \mathbf{x}_i), \tag{2}$$

where $\mathbf{x}_i$ represents the target clean reference. Thus, the resultant Θ^* becomes independent of the noise level σ, because f_{Θ^*} is able to minimize the loss $\mathcal{L}$ in Eqn. (2) for arbitrary σ. However, Θ^* still depends on the noise level of the target image $\mathbf{x}_i$, which indicates that Eqn. (2) is equivalent to

$$\Theta^* = \arg\min_{\Theta} \sum_i \mathcal{L}(f(\mathbf{y}_i, \sigma_{\mathbf{y}_i}; \Theta(\sigma_{\mathbf{x}_i})), \mathbf{x}_i)|_{\sigma_{\mathbf{x}_i}=0}, \tag{3}$$

where $\sigma_{\mathbf{y}_i}$ and $\sigma_{\mathbf{x}_i}$ represent the noise level of the source and the target images, respectively. If we follow the transformation in FFDNet [36] and take $\sigma_{\mathbf{x}_i}$ as an input variable of the mapping function, we can obtain a general form of Eqn. (3), as follows:

$$\Theta^* = \arg\min_{\Theta} \sum_i \mathcal{L}(f(\mathbf{y}_i, \sigma_{\mathbf{y}_i}, \sigma_{\mathbf{z}_i}; \Theta), \mathbf{z}_i), \tag{4}$$

where the target image $\mathbf{z}_i$ can be a corrupted image with the noise level $\sigma_{\mathbf{z}_i}$. Eqn. (4) implies that the resultant model becomes a noise transference network, which can be trained without any clean reference.

In addition, we mainly consider the realistic noise in this work, and thus the noise level of the input observation $\sigma_{\mathbf{y}_i}$ is not available for training. Therefore, we adopt the "residual-on-residual" [3] architecture to establish NTGAN, and omit the source noise level $\sigma_{\mathbf{y}_i}$ in Eqn. (4), as follows:

$$\Theta^* = \arg\min_{\Theta} \sum_i \mathcal{L}(f(\mathbf{y}_i, \sigma_{\mathbf{z}_i}; \Theta), \mathbf{z}_i). \tag{5}$$

Eqn. (5) indicates that NTGAN is a blind denoiser, which does not require *a priori* knowledge on $\sigma_{\mathbf{y}_i}$ for noise removal. It is worth noting that the newly introduced input variable $\sigma_{\mathbf{z}_i}$ controls the noise level we expect NTGAN to synthesize, which is manually set during the noise generation process, and thus it is available for training. Moreover, during the denoising inference, the target noise level is set to zero as follows:

$$\hat{\mathbf{x}} = f(\mathbf{y}, \sigma_{\mathbf{z}}; \Theta^*)|_{\sigma_{\mathbf{z}}=0}, \tag{6}$$

because we expect NTGAN to generate a noise-free ($\sigma_{\mathbf{z}} = 0$) image. On the other hand, the noise level of a real-world noisy image cannot be simply described by a single value $\sigma_{\mathbf{z}}$. Therefore, we follow the settings in CBDNet [11], and establish the noise level map $\mathbf{M}_{\mathbf{z}}$ to replace $\sigma_{\mathbf{z}}$ in Eqns. (5) and (6), which consists of the pixel-wise noise standard deviations in the target observation $\mathbf{z}$.

In summary, NTGAN is a conditional generator for different noise levels. It takes a noisy observation $\mathbf{y}$ and a target noise level map $\mathbf{M}_{\mathbf{z}}$ as inputs, and produces a new observation $\mathbf{z}$ with the guiding (expected) noise level $\mathbf{M}_{\mathbf{z}}$.

3.2 Pseudo target noise level map

As introduced in Sec. 3.1, NTGAN requires the noise level maps of the target noisy images for training. However, such a paired set, $[\mathbf{y}, (\mathbf{z}, \mathbf{M}_{\mathbf{z}})]$, is hard to obtain when only noisy samples are available. As the noise level of $\mathbf{y}$ is not available, leading to the difficulty in obtaining true $\mathbf{M}_{\mathbf{z}}$. Therefore, we further propose a strategy for generating pseudo target noise level maps when training with the real-world observations. As the noise residing in real photographs is relatively weak compared to the image content, we can synthesize the pairs of observations by adding the synthetic noise to a real photograph $\mathbf{y}$ as follows:

$$\mathbf{z} = \mathbf{y} + \mathbf{n}, \tag{7}$$

where $\mathbf{n}$ is the synthetic noise from a noise generation model. Thus, the target noise level map $\mathbf{M}_{\mathbf{z}}$ can be approximated as the noise level of $\mathbf{n}$, which can be obtained from the noise generation process. NTGAN aims to learn the transference between the paired observations with different noise intensities. Therefore, the noise residing in $\mathbf{y}$ and $\mathbf{z}$ should follow the same distribution. We assume that the realistic noise follows a heterogeneous Gaussian distribution, and so adopt the noise generation model from [21], which considers both the signal-independent and signal-dependent components in $\mathbf{n}$. In addition, we also apply the in-camera pipeline in [11] to the noise generation model to further narrow the domain gap between the noise in $\mathbf{y}$ and $\mathbf{z}$. Thus, the resultant noise generation model is formulated as follows:

$$\begin{aligned} \mathbf{n}(\mathbf{y}, \mathbf{M}_{\mathbf{z}}) &= f_{\text{BPD}}(f_{\text{crf}}(\mathbf{L} + \mathbf{n}_s(\mathbf{L}) + \mathbf{n}_c)) - f_{\text{BPD}}(f_{\text{crf}}(\mathbf{L})), \\ &\text{with } \mathbf{L} = f_{\text{icrf}}(\mathbf{y}), \end{aligned} \tag{8}$$
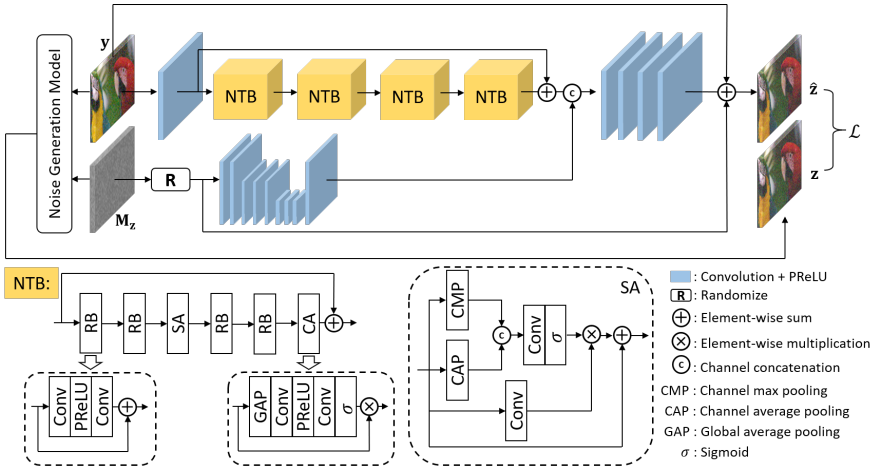
Figure 1: The proposed NTGAN for blind unsupervised image denoising. The last convolutional layers in both branches are not followed by PReLU. NTB: noise transference block, RB: residual block. SA: spatial attention unit. CA: channel attention unit.

where f_{crf} and f_{icrf} represent the camera response function and the inverse camera response function, respectively. Specifically, f_{icrf} transforms the original image $\mathbf{y}$ to obtain an irradiance plane $\mathbf{L}$. Moreover, $\mathbf{n}_s$ and $\mathbf{n}_c$ account for the noise components that are dependent and independent of the signal $\mathbf{y}$, respectively. Therefore, the pseudo target noise level map $\mathbf{M}_{\mathbf{z}}$ is defined as follows:

$$\mathbf{M}_{\mathbf{z}}^2 = \mathbf{L} \cdot \sigma_s^2 + \sigma_c^2, \tag{9}$$

where $\mathbf{L} \cdot \sigma_s^2$ and σ_c^2 are the noise variance of $\mathbf{n}_s$ and $\mathbf{n}_c$, respectively. f_{BPD} in Eqn. (8) represents the function considering the Bayer patterning and demosaicing, which was proposed in [15] for simulating the spatially correlated noise. This noise generation model guarantees the statistics consistency of the noise in $\mathbf{y}$ and $\mathbf{z}$, because the resultant noise in both $\mathbf{y}$ and $\mathbf{z}$ is heterogeneous Gaussian distributed with the dependency on the clean signal. With the proposed strategy for the pseudo noise level map, the training pairs can be obtained from the noisy observations by using Eqns. (7), (8), and (9) to form the training set as $\{[\mathbf{y}_1, (\mathbf{M}_{\mathbf{z}_1}, \mathbf{z}_1)]$, $[\mathbf{y}_2, (\mathbf{M}_{\mathbf{z}_2}, \mathbf{z}_2)], \ldots, [\mathbf{y}_n, (\mathbf{M}_{\mathbf{z}_n}, \mathbf{z}_n)]\}$.

3.3 Framework

As introduced in Sec. 3.1, we aim to introduce the noise level of the target image as an input variable of the model, so that the resultant network can be a conditional noise generator, controlled by $\mathbf{M}_{\mathbf{z}}$. In addition, the "residual-on-residual" structure is required to omit the supervision from $\sigma_{\mathbf{y}}$. Therefore, we establish the network as shown in Figure 1. Specifically, NTGAN consists of an image-encoding stream and a noise-level-encoding stream, which are presented as the top and the bottom branches in Figure 1, respectively. The image-encoding branch consists of four noise transference blocks (NTB), each of which contains a spatial attention (SA) unit, a channel attention (CA) unit and four residual blocks (RB). The two attention units follow the designs in [8] and [37], in order to focus on the important regions in both the spatial and channel dimensions. Specifically, in the SA unit, an input feature map

$F \in \mathbb{R}^{c \times h \times w}$ is first independently compressed by average pooling and max pooling in the channel dimension. The two compact feature maps are concatenated, and fed to a convolution layer with the sigmoid function to form a heat-map $H \in \mathbb{R}^{1 \times h \times w}$. We use H to rescale F along the spatial dimension for spatial attention. Similarly in CA, the global average pooling is first applied to F to reduce the feature size to $v \in \mathbb{R}^{c \times 1 \times 1}$. We generate the channel heat-vector by passing v to a non-linear mapping module with the sigmoid function. The produced heat-vector is used to rescale the different channels in F to achieve channel attention. All RB, NTB, and the global streams employ the residual connections, which inherits the 'residual-on-residual'' structure in [3] to achieve noise transference without knowing the source noise level. The noise-level-encoding branch consists of convolutional layers, average pooling layers, and a randomization block. The randomization is defined as the element-wise multiplication with a set of standard Gaussian distributed random variable $r \sim \mathcal{N}(0,1)$, which follows the design in [16]. The features from the two streams are fused by the channel concatenation. Finally, we reconstruct the target observation $\hat{z}$,based on the fused features, to mislead a patch-based discriminator.

To learn the noise transference task, we consider two loss terms in the objective function for training. NTGAN aims to produce a noisy observation with the expected noise level. Thus, the produced noisy image aims to fool the discriminator, while maintaining the image content. Therefore, the training objective is formulated as follows:

$$\mathcal{L} = \mathcal{L}_{\text{GAN}} + \lambda \mathcal{L}_{\text{rec}}, \tag{10}$$

where $\mathcal{L}_{\text{GAN}}$ and $\mathcal{L}_{\text{rec}}$ are the loss of the discriminator and the reconstruction, respectively, and λ is a hyperparameter controlling the trade-off between these two loss terms. Specifically, $\mathcal{L}_{\text{rec}}$ is computed as follows:

$$\mathcal{L}_{\text{rec}} = \frac{1}{n} \sum_{i=1}^{n} ||\hat{z}_i - z_i||_2^2, \tag{11}$$

where $\hat{z}$ denotes the reconstructed observation from NTGAN, and z is the target observation based on the noise generation model. $\mathcal{L}_{\text{GAN}}$ is defined as follows:

$$\mathcal{L}_{\text{GAN}} = \mathbb{E}_{M_z,z}[\log(D(M_z, z))] + \mathbb{E}_{M_z,y}[\log(1 - D(M_z, G(M_z, y)))], \tag{12}$$

where D denotes the patch-based discriminator [11], and G denotes the generator, which consists of all the modules in NTGAN for producing $\hat{z}$.

4 Experiments

4.1 Implementation details

NTGAN is an unsupervised denoiser, and thus it only requires noisy observations for training. In this work, we mainly consider the realistic noise removal task. Therefore, we evaluate NTGAN on two commonly used real-world denoising benchmarks, i.e. the default 15 cropped noisy images in the Cross-Channel dataset (CC15) [23] and the Darmstadt Noise dataset (DND) [26]. To make a fair comparison with the other denoising methods, we use the cropped images of size 512×512 from SIDD [1], PolyU [32], and RENOIR [2] to form the training set. In the training phase, we randomly crop $8,000$ image patches with size 64×64 from the noisy images in the training set, and synthesize the training pairs using

| (a) Noisy | (b) BM3D | (c) NI | (d) DnCNN+ | (e) CBDNet | (f) VDNet |

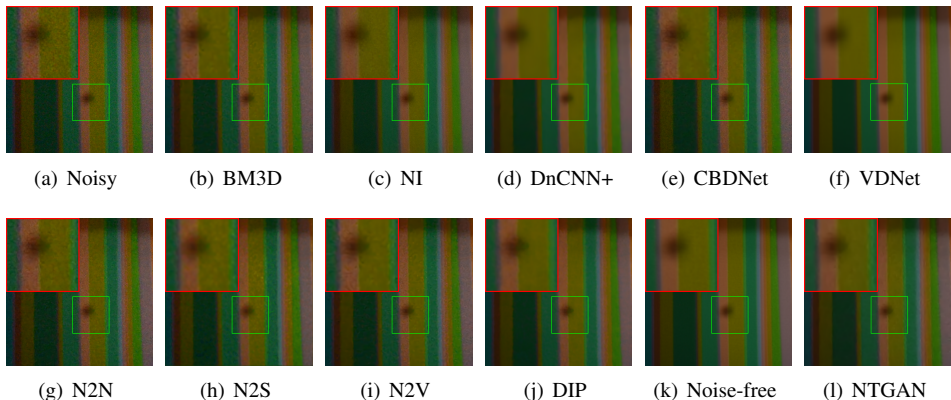| (g) N2N | (h) N2S | (i) N2V | (j) DIP | (k) Noise-free | (l) NTGAN |

Figure 2: Evaluation on the perceptual qualities of a real-world noisy image from CC15, restored by different methods.

Eqns. (7), (8), and (9). In addition, the noise variance σ_s and σ_c in Eqn. (9) are uniformly sampled from the range $(0, 0.06]$ and $(0, 0.03]$, respectively. Random rotation and mirroring are applied to the patches for augmentation. It is worth noting that, although the nearly noise-free images are available in SSID , PolyU, and RENOIR, we only use their noisy observations for training. In the testing phase, we feed the noisy observations from DND and CC15 to the trained NTGAN, and set all the elements in $\mathbf{M_z}$ to zero, as we expect NTGAN to generate the noise-free ($\mathbf{M_z} = 0$) images.

We implement NTGAN based on PyTorch [25]. All the convolutional layers consists of 3×3 kernels with padding 1 and stride 1 except for those in the SA units and before the sigmoid function, where the kernel size is 1×1 with padding 0. The channel number of NTB is fixed to 64 except for that in the CA units, where 4 convolutional filters are used. We employ Adam [12] to optimize the objective function defined in Eqn. (10) with the trade-off controller λ empirically set to 0.3. We train the network for 10^6 iterations on two Nvidia GEFORCE GTX 1080 Ti GPUs, with batch size 128, and the learning rate is set to 10^{-4} and halved at the 5×10^5th iteration.

4.2 Evaluation on realistic noise removal

To evaluate the denoising performance of NTGAN on realistic noise, we applied it to the real noisy photographs in CC15 and DND.

Type	Traditional method		Supervised network			Unsupervised network				
Method	BM3D	NI	DnCNN+	CBDNet	VDNet	N2N	DIP	N2S	N2V	Ours
PSNR	35.19	35.33	35.40	36.44	35.84	35.32	35.69	35.38	35.27	**37.33**
SSIM	0.9063	0.9212	0.9115	0.9460	0.9414	0.9160	0.9259	0.9204	0.9158	**0.9476**

Table 1: The quantitative results on CC15. The best results are highlighted in **bold**.

The Cross-Channel dataset (CC15) [23] consists of eleven noisy images with static scenes captured by three different cameras. The nearly noise-free ground-truth is obtained by taking the average over the 500 shots with the same camera settings of each observation. We compare our proposed method with the traditional methods, i.e. BM3D [8] and NI [24], the supervised CNN-based methods, i.e. DnCNN+ [35], CBDNet [10], and VDNet [34],

and the state-of-the-art unsupervised methods, i.e. N2N [17], DIP [30], N2S [4] and N2V [14], in terms of the Peak Signal-to-Noise Ratio (PSNR) and the structural similarity (SSIM) index. To make a fair comparison, we adopt the default settings or the pre-trained models provided by the original authors of those compared methods. The quantitative results are listed in Table 1. It can be observed that the proposed NTGAN achieves the best performance on CC15. NTGAN outperforms the supervised networks, because the proposed noise transference strategy can be regarded as a special data augmentation method for the denoising task. NTGAN learns the mapping from one noisy observation to multiple observations with the reliable guidance from $\mathbf{M_z}$, which helps the network to better memorize the noise distributions. Compared with the other unsupervised networks, NTGAN can generally obtain a PSNR gain of over 1dB, which results from the utilization of the pseudo noise level maps in guiding the learning for different noise distributions. The qualitative comparison is presented in Figure 2. It can be observed that the traditional and the unsupervised competitors tend to retain some noise in the restored image, while DnCNN+ and VDNet oversmooth the image. Although CBDNet generally produces comparable visual results to NTGAN, it creates some distortions on the edges, as shown in the upscaled patches.

Type	Traditional method			Supervised network						Unsupervised network
Method	BM3D	WNNM	NI	DnCNN+	CBDNet	RIDNet	VDNet	GCBD	ADGAN	Ours
PSNR	34.51	34.67	35.11	37.90	38.05	39.23	39.38	37.72	38.13	**39.42**
SSIM	0.8507	0.8646	0.8778	0.9430	0.9421	0.9526	0.9518	0.9408	**0.9580**	0.9560

Table 2: The quantitative results on DND. The best results are highlighted in **bold**.

The Darmstadt Noise dataset (DND) [26] is a recently proposed large-scale real-world denoising dataset, which consists of $1,000$ paired noisy and nearly noise-free image patches. The nearly noise-free images are not publicly available, and thus the results can only be obtained through their online submission system. As the above-mentioned unsupervised methods were not evaluated on DND, we just compare NTGAN with the supervised networks and the traditional methods. In addition, GANs have been widely studied and employed for image denoising, as shown in the literature. Thus, we also compare the proposed NTGAN with some state-of-the-art GAN-based denoisers on DND. GAN-based denoisers can be generally categorized into two groups, i.e. using the generator to restore the noisy images, such as ADGAN [19] and using GAN to model the noise generation process for providing training samples, such as GCBD [6]. The quantitative results are tabulated in Table 2. It can be seen that NTGAN outperforms the other supervised and traditional competitors, which demonstrates its effectiveness in learning the noise transference task as the special augmentation approach for noise removal. Compared to the GAN-based methods, NTGAN gains a PSNR improvement of over 1dB on DND. This is because the proposed NTGAN utilizes the reliable pseudo supervision from the target noise level in training, which facilitates the learning for different noise distributions. Considering that the clean reference is not required, the proposed strategy has greater potential in practical applications.

4.3   Evaluation on noise transference

To evaluate the performance of NTGAN for noise transference, we consider two cases, i.e. transferring the synthetic AWGN and transferring the realistic noise. As NTGAN synthesizes the target observation based on the guiding noise level map, we employ the pre-trained noise level estimation sub-network in CBDNet [10] to extract the target $\mathbf{M_z}$ from a DND image for guiding the realistic noise transference task. The results are presented in Fig. 3. For

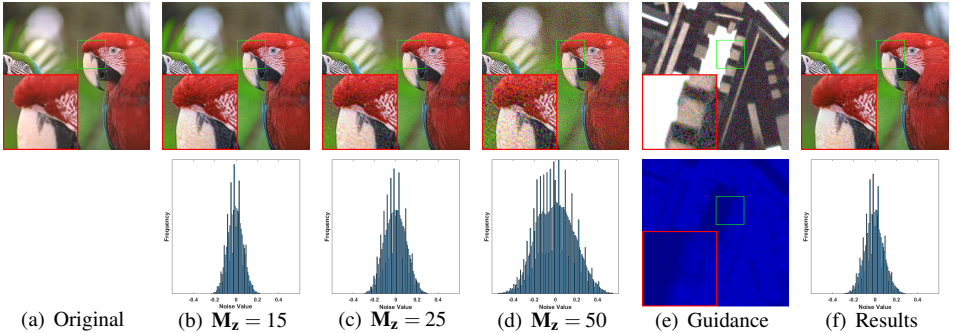| (a) Original | (b) $\mathbf{M_z} = 15$ | (c) $\mathbf{M_z} = 25$ | (d) $\mathbf{M_z} = 50$ | (e) Guidance | (f) Results |

Figure 3: The evaluation on noise transference. In the first row, (b)-(d) are the synthetic noisy images based on the uniform noise level maps, (e) presents the guiding image selected from DND, and (f) presents the transference result based on the noise level map extracted from (e). In the second row, (b)-(d) and (f) present the noise distribution in the synthetic images, and (e) presents the estimated noise level map of the guiding image.

simplicity, we set the input observation as a clean image. It can be observed from Figs. 3(b)-3(d) that when $\mathbf{M_z}$ increases, NTGAN introduces stronger noise into the synthetic image. More importantly, when guided by the realistic noise level map in Fig. 3(e), the noise in the transferred image follows the same pattern as that in the guiding image, which demonstrates the effectiveness of NTGAN in memorizing the signal-dependent noise.

4.4 Ablation study

$\mathcal{L}_{GAN}$	✓	✓	✓	✓	✓		✓	✓	✓
RoR		✓	✓	✓		✓	✓	✓	✓
CA			✓		✓	✓	✓	✓	✓
SA				✓	✓	✓	✓	✓	✓
N2C								✓	
N2N									✓
CC15	35.50	36.41	36.73	36.67	35.88	37.25	37.33	37.21	36.52
DND	37.68	38.06	38.40	38.58	37.87	39.33	39.42	39.36	38.49

Table 3: The ablation study on DND and CC15 in term of PSNR. $\mathcal{L}_{GAN}$: GAN loss. RoR: residual-on-residual structure. CA: channel attention unit. SA: spatial attention unit. N2C: noise-to-clean manner. N2N: noise-to-noise manner.

To provide a comprehensive analysis on NTGAN, we perform ablation studies to investigate the effectiveness of the different designs in NTGAN. Specifically, we consider the effect of the GAN loss, the residual-on-residual structure (RoR), the channel attention unit, and the spatial attention unit in NTGAN. In addition, we also consider training the network in a noise-to-clean (N2C) manner and a noise-to-noise (N2N) manner. Specifically, N2C uses the clean references as the target images for training, while N2N detaches the noise-level-encoding branch and fits the noisy observations without the guidance from the target noise level maps. To make a fair comparison, we enlarge the kernel size and the network depth of the models without CA, SA, or the noise-level-encoding branch, in order to make their model capacity equal to, or larger than, the original NTGAN. It can be observed from Table 3 that the denoising performance is slightly degraded when omitting $\mathcal{L}_{GAN}$. Moreover, we

observe that the network, trained without $\mathcal{L}_{\text{GAN}}$, fails to perform accurate noise transference. If we increase the value of the target noise level, the network produces more blurred images. Both attention units contribute to a PSNR improvement of about 1dB and 0.5dB on DND and CC15, respectively. The residual-on-residual (RoR) structure plays an important role in blind denoising. Without RoR, the performance is dramatically degraded, because the network cannot flexibly deal with the input observations with different noise levels. When N2C is applied, the results are slightly degraded, compared to the noise transference strategy. This is because noise transference, as introduced in the previous section, can serve as a special augmentation approach for the denoising task. Moreover, when the model is trained in an N2N manner, the performance is dramatically degraded. The proposed noise transference strategy utilizes the guidance from the noise intensity at the different positions of the observation in training, which facilitates the learning for the distribution of the spatially variant noise, and consequently enhances the model generalization on real photographs. Therefore, the proposed noise transference strategy is beneficial to the denoising performance.

5   Conclusion

In this paper, we have proposed a novel noise transference generative adversarial network, namely NTGAN, for blindly removing the realistic noise residing in photographs. Our proposed method treats the noise transference task as the general form of the noise reduction task, which enables it to obtain the denoising ability by observing the corrupted data only. NTGAN learns the noise transference task based on the proposed pseudo noise level strategy, which provides the reliable guidance for learning different noise distributions. We have evaluated NTGAN on two widely used real-world denoising benchmarks. The experimental results have showed that our proposed method achieves promising performance on real photographs, making it a potential solution to practical denoising problems.

References

[1] A. Abdelhamed, S. Lin, and M. S. Brown. A high-quality denoising dataset for smartphone cameras. In *CVPR*, pages 1692–1700, June 2018.

[2] 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.

[3] Saeed Anwar and Nick Barnes. Real image denoising with feature attention. In *ICCV*, October 2019.

[4] Joshua Batson and Loïc Royer. Noise2self: Blind denoising by self-supervision. *arXiv*, 2019.

[5] Guangyong Chen, Fengyuan Zhu, and Pheng Ann Heng. An efficient statistical method for image noise level estimation. In *ICCV*, pages 477–485, 2015.

[6] J. Chen, J. Chen, H. Chao, and M. Yang. Image blind denoising with generative adversarial network based noise modeling. In *CVPR*, pages 3155–3164, 2018.

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

[8] Kostadin Dabov, Alessandro Foi, Vladimir Katkovnik, and Karen O. Egiazarian. Image denoising by sparse 3-d transform-domain collaborative filtering. *IEEE Transactions on Image Processing*, 16:2080–2095, 2007.

[9] Shuhang Gu, Lei Zhang, Wangmeng Zuo, and Xiangchu Feng. Weighted nuclear norm minimization with application to image denoising. In *CVPR*, pages 2862–2869, 2014.

[10] Shi Guo, Zifei Yan, Kai Zhang, Wangmeng Zuo, and Lei Zhang. Toward convolutional blind denoising of real photographs. In *CVPR*, pages 1712–1722, June 2019.

[11] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A. Efros. Image-to-image translation with conditional adversarial networks. In *CVPR*, pages 5967–5976, 2017.

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

[13] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton. Imagenet classification with deep convolutional neural networks. *Commun. ACM*, 60:84–90, 2012.

[14] Alexander Krull, Tim-Oliver Buchholz, and Florian Jug. Noise2void - learning denoising from single noisy images. In *CVPR*, June 2019.

[15] Claude A Laroche and Mark A Prescott. Apparatus and method for adaptively interpolating a full color image utilizing chrominance gradients, Dec. 1994. US Patent 5,373,322.

[16] Anders Boesen Lindbo Larsen, Søren Kaae Sønderby, Hugo Larochelle, and Ole Winther. Autoencoding beyond pixels using a learned similarity metric. In *ICML*, volume 48, pages 1558–1566, 20–22 Jun 2016.

[17] Jaakko Lehtinen, Jacob Munkberg, Jon Hasselgren, Samuli Laine, Tero Karras, Miika Aittala, and Timo Aila. Noise2Noise: Learning image restoration without clean data. In *ICML*, volume 80, pages 2965–2974, July 2018.

[18] Jun Li, Hongfu Liu, and Yun Fu. Predictive coding machine for compressed sensing and image denoising. In *AAAI*, 2018.

[19] Kai Lin, Thomas H. Li, Shan Liu, and Ge Li. Real photographs denoising with noise domain adaptation and attentive generative adversarial network. In *CVPR Workshops*, June 2019.

[20] Ce Liu, William T. Freeman, Richard Szeliski, and Sing Bing Kang. Noise estimation from a single image. In *CVPR*, pages 901–908, 2006.

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

[22] Mehdi Mirza and Simon Osindero. Conditional generative adversarial nets. *ArXiv*, abs/1411.1784, 2014.

[23] Seonghyeon Nam, Youngbae Hwang, Yasuyuki Matsushita, and Seon Joo Kim. A holistic approach to cross-channel image noise modeling and its application to image denoising. In *CVPR*, pages 1683–1691, 2016.

[24] Neatlab and ABSoft. Neat image. *http://ni.neatvideo.com/home*, 2001.

[25] Adam Paszke, Sam Gross, Soumith Chintala, Gregory Chanan, Edward Yang, Zachary DeVito, Zeming Lin, Alban Desmaison, Luca Antiga, and Adam Lerer. Automatic differentiation in PyTorch. In *NIPS Autodiff Workshop*, 2017.

[26] Tobias Plotz and Stefan Roth. Benchmarking denoising algorithms with real photographs. In *CVPR*, pages 2750–2759, 2017.

[27] Javier Portilla. Full blind denoising through noise covariance estimation using gaussian scale mixtures in the wavelet domain. In *ICIP*, volume 2, pages 1217–1220 Vol.2, 2004.

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

[29] Shaoqing Ren, Kaiming He, Ross B. Girshick, and Jian Sun. Faster r-cnn: Towards real-time object detection with region proposal networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39:1137–1149, 2015.

[30] Dmitry Ulyanov, Andrea Vedaldi, and Victor S. Lempitsky. Deep image prior. In *CVPR*, pages 9446–9454, 2017.

[31] Jun Xu, Lei Zhang, Dongxiao Zhang, and Xiangchu Feng. Multi-channel weighted nuclear norm minimization for real color image denoising. In *ICCV*, pages 1105–1113, 2017.

[32] Jun Xu, Hui Li, Zhetong Liang, David Zhang, and Lei Zhang. Real-world noisy image denoising: A new benchmark. *arXiv*, 04 2018.

[33] Jun Xu, Lei Zhang, and Dongxiao Zhang. A trilateral weighted sparse coding scheme for real-world image denoising. In *ECCV*, 2018.

[34] Zongsheng Yue, Hongwei Yong, Qian Zhao, Deyu Meng, and Lei Zhang. Variational denoising network: Toward blind noise modeling and removal. In *NIPS*, pages 1690–1701, 2019.

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

[36] Kai Zhang, Wangmeng Zuo, and Lei Zhang. Ffdnet: Toward a fast and flexible solution for cnn-based image denoising. *IEEE Transactions on Image Processing*, 27:4608–4622, 2017.

[37] Yulun Zhang, Kunpeng Li, Kai Li, Bineng Zhong, and Yun Fu. Residual non-local attention networks for image restoration. In *ICLR*, 2019.

[38] R. Zhao, K. Lam, and D. P. K. Lun. Enhancement of a cnn-based denoiser based on spatial and spectral analysis. In *ICIP*, pages 1124–1128, 2019.