Noise Distribution Adaptive Self-Supervised Image Denoising using Tweedie Distribution and Score Matching

Kwanyoung Kim[1]   Taesung Kwon[2]   Jong Chul Ye[1,2]

[1]Kim Jaechul Graduate School of AI, [2] Department of Bio and Brain Engineering, KAIST

{cubeyoung, star.kwon, jong.ye}@kaist.ac.kr
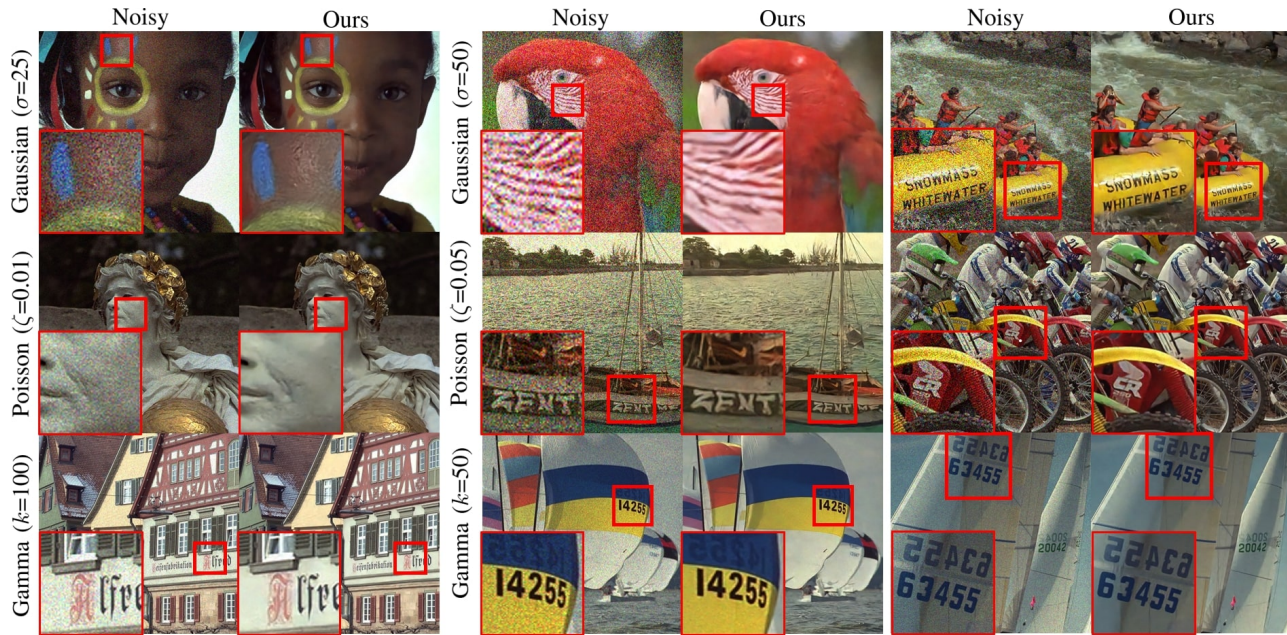
Figure 1. Our image denoising results in Kodak dataset. Top: Gaussian noise (σ =25, $\sigma = 50$), Middle: Poisson noise ($\zeta = 0.01, 0.05$), Bottom: Gamma noise ($k = 100, 50$).

Abstract

Tweedie distributions are a special case of exponential dispersion models, which are often used in classical statistics as distributions for generalized linear models. Here, we show that Tweedie distributions also play key roles in modern deep learning era, leading to a distribution adaptive self-supervised image denoising formula without clean reference images. Specifically, by combining with the recent Noise2Score self-supervised image denoising approach and the saddle point approximation of Tweedie distribution, we provide a general closed-form denoising formula that can be used for large classes of noise distributions without ever knowing the underlying noise distribution. Similar to the original Noise2Score, the new approach is composed of two successive steps: score matching using perturbed noisy images, followed by a closed form image denoising formula via distribution-independent Tweedie's formula. In addition, we reveal a systematic algorithm to estimate the noise model and noise parameters for a given noisy image data set. Through extensive experiments, we demonstrate that the proposed method can accurately estimate noise models and parameters, and provide the state-of-the-art self-supervised image denoising performance in the benchmark dataset and real-world dataset.

1. Introduction

Image denoising is a fundamental problem in low-level computer vision problems. Nowadays, typical supervised learning approaches easily outperform classical denoising algorithms such as Block-Matching and 3D filtering (BM3D) [2] and Weighted Nuclear Norm Minimization (WNNM) [5]. Nonetheless, the supervised approaches are
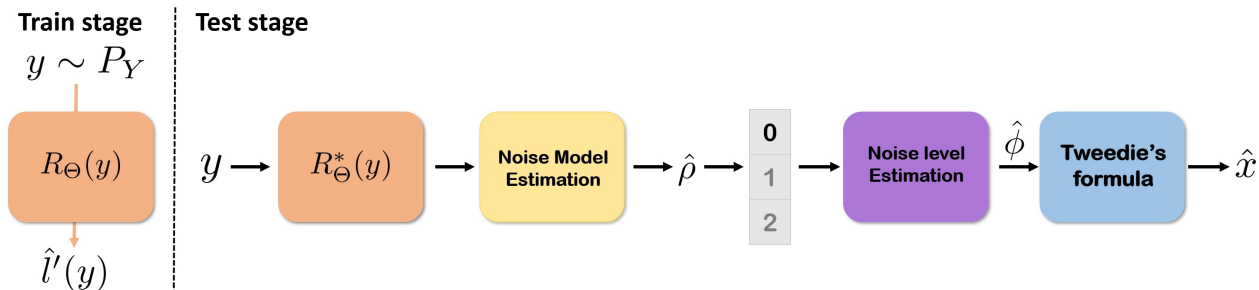
Figure 2. Overall reconstruction flow of the proposed method, where the first step is the estimation of the score function $\hat{l}'$ by training a neural network R_Θ, which is followed by the estimation of noise model and noise level to obtain final denoised results. During the training procedure, input images y that belong to the distribution of train set P_Y, are sampled.

not practical in many real-world applications as they require a large number of matched clean images for training.

To address this issue, researchers have proposed various forms of self-supervised learning approaches trained with ingenious forms of loss functions that are not associated with clean reference images [1,6,9,12,14,21]. Specifically, these approaches have focused on designing loss functions to prevent from learning identity mapping, and can be categorized into two classes: 1) one with generating altered target images from noisy input images [1, 6, 12, 14], and 2) the other by adding regularization terms from Stein's Unbiased Risk Estimation (SURE) [9, 21].

Although these algorithms appear seemingly different, a recent Noise2Score [10] revealed that the procedure of generating altered target images or SURE-based regularization term is closely related to the score matching [7], and there exists a Bayes optimal denoising formula in terms of score function for any exponential family distributions. Unfortunately, Noise2Score requires a prior knowledge of the noise distribution, so when the underlying noise distribution is unknown, Noise2Score provide a sub-optimal performance.

One of the most important contributions of this paper is, therefore, a novel discovery that the classical Tweedie distribution can provide a "magic" recipe that can be used for a large class of noise distribution even without knowing the distribution. Specifically, Tweedie distribution can be synergistically combined with Noise2Score to provide an explicit de-noising formulation and an algorithm for estimating the underlying noise model and parameters. In particular, inspired by the fact that various exponential family distributions like Gaussian, Gamma, Poisson, etc. can be described by saddle point approximation of the Tweedie distribution by simply changing one parameter, we provide a universal noise removal formula that can be used for a large class exponential family distributions without prior knowledge of the noise model. Furthermore, by assuming that slightly perturbed noisy image may produce similar denoising results, we provide a systematic algorithm that

can estimate the noise type and associate parameters for any given images. In spite of the blind nature of the algorithm, experimental results demonstrated that our method outperforms other self-supervised image denoising methods that are trained with prior knowledge of noise distributions.

Our contribution can be summarized as follows.

- We provide a general closed-form denoising formula for large classes of noise distributions by combining Noise2Score approach and the saddle point approximation of Tweedie distribution.

- We propose an algorithm to estimate the noise model and noise parameter for given noisy images. In particular, the proposed noise estimation algorithm significantly improves the performance and boosts the inference speed compared to the original Noise2Score [10].

- We show that the proposed method produces the state-of-the-art performance amongst various self-supervised image denoising algorithms in the benchmark dataset and real-environment dataset.

2. Related Works

2.1. Self-supervised image denoising

Recently, self-supervised image denoising methods using only noisy images have been widely explored. Noise2Noise [14] (N2N) was proposed to train a neural network by minimizing L_2 distance between the noisy image and another noisy realization of the same source image. When additional noisy versions of the same image are not available, Noise2Void (N2V) [12], Noise2Self (N2S) [1] adopted the blind spot network to avoid learning the identity noisy images. Other types of mask-based blind-spot network methods have been explored, which include Self2Self [20], Noise2Same [25], Laine19 [13], etc. Instead of using blind-spot methods, Neighbor2Neighbor (Nei2Nei) [6] generate training image pairs by using the random neighbor sub-sampler method. Noisier2Noise [18]

samples doubly noisy images by adding synthetic noises on the single noisy input. Additionally, Soltanayev et al [21] proposed a training scheme using Stein's unbiased risk estimator (SURE) to denoise additive Gaussian noise. By extending this idea, in [9], the authors proposed a loss function using Poisson Unbiased risk estimator (PURE) to train deep neural network for Poisson noises.

However, these methods have been mostly designed either by distribution independent heuristics or for specific noise models, and we are not aware of any general methods that can be used for a large class of noise distributions by estimating the noise model and parameters.

2.2. Noise2Score

Recently, Noise2Score [10] was proposed to estimate clean images by Tweedie's formula given the learned score function of noisy data. Specifically, Noise2Score [10] consists of two steps: score function estimation using noisy images, which is followed by Tweedie's formula to estimate the clean images. To learn the score function from noisy data y, Noise2Score employes the amortized residual DAE (AR-DAE) [15], which is a stabilized implementation of denoising autoencoder (DAE) [24]. In fact, this procedure is closely related to the masking procedure in Noise2Void, Noise2Self, Noiser2Noise, etc. Specifically, Noise2Score tried to minimize the following loss:

$$\underset{\Theta}{\arg\min} = \underset{y \sim P_Y}{\mathbb{E}} \|u + \sigma_a R_\Theta(y + \sigma_a u)\|^2 \qquad (1)$$

where R_Θ is the score model parameterized by Θ, $u \sim \mathcal{N}(0, I)$, and $\sigma_a \sim \mathcal{N}(0, \delta^2)$. σ_a is perturbed noise which gradually decrease with an annealing schedule. Minimizing Equation (1) yields the network R_Θ^* which could directly estimate the score function of noisy data, $R_\Theta^* = \nabla_y \log p(y) = l'(y)$.

Once the score function $l'(y)$ for the noisy image is learned, Noise2Score provides an explicit formula for the clean image using Tweedie's formula. More specifically, consider an exponential family of probability distributions:

$$p(y|\eta) = p_0(y) \exp \left(\eta^\top T(y) - A(\eta) \right) \qquad (2)$$

where the superscript $^\top$ denotes the transpose operation. Here, η is a canonical (vector) parameter of the family, $T(y)$ is a (vector) function of y, $A(\eta)$ is the cumulant generating function which makes $p(y|\eta)$ integrate to 1, and $p_0(y)$ the density up to a scale factor when $\eta = 0$. Then, Tweedie's formula [4] shows that the posterior estimate of the canonical parameter $\hat{\eta}$ should satisfy the following equation:

$$\hat{\eta}^\top T'(y) = -l_0'(y) + l'(y) \qquad (3)$$

where $l_0'(y) := \nabla_y \log p_0(y)$ are score functions, and $T'(y) = \nabla_y T(y)$ [10].

3. Main Contribution

In practice, when noisy images are collected, it is difficult to assume the underlying noise models. For example, low dose X-ray CT images are contaminated with a lot of noises, but it is not clear whether it is from Gaussian noise, Poisson noises, or mixed distribution. Usually, a specific noise model is assumed for noise removal, but when the assumed noise model is incorrect, there exists an unavoidable bias in the estimated images. The situation is same in Noise2Score, as a specific form of Tweedie's formula is used depending on the assumed noise model. Therefore, a fundamental question for image denoising is whether there is a universal algorithm that can be used for various noise models without losing its optimality.

In this section, we answer this question affirmatively. Specifically, we show that the classical Tweedie distribution model is a versatile tool that can be combined with Noise2Score to provide a near optimal solution for various noise distribution without even knowing the underlying noise model.

3.1. Tweedie distribution

The Tweedie distributions are a family of probability distributions which include the purely continuous normal, gamma and Inverse Gaussian distributions, the purely discrete scaled Poisson distribution, and the class of compound Poisson–gamma distributions [8]. Specifically, for a random variable y which follows an exponential dispersion model, the Tweedie density function is given by

$$p(y; \mu, \phi) = b_\rho(y, \phi) \exp \left(\frac{-d(y, \mu)}{2\phi} \right), \qquad (4)$$

where μ is the mean, $\phi > 0$ is the dispersion parameter which is related to noise level, $d(y, \mu)$ is the unit deviance for $\rho \neq 0, 1, 2$:

$$d(y, \mu) = 2 \left(\frac{y^{2-\rho}}{(1-\rho)(2-\rho)} - \frac{y\mu^{1-\rho}}{1-\rho} + \frac{\mu^{2-\rho}}{2-\rho} \right). \qquad (5)$$

Tweedie family densities are characterized by power variance functions of the form $V[\mu] = \phi\mu^\rho$, where $\rho \in (-\infty, 0] \cup [1, \infty)$ is the index determining the distribution [8]. This has a closed form expression for special cases, such as the normal distribution ($\rho = 0$), Poisson ($\rho = 1$), gamma ($\rho = 2$) and inverse Gaussian ($\rho = 3$) distributions (see Table 1). However, the function $b_\rho(y, \phi)$ cannot be written in closed form except for special cases.

Thus, the saddle point approximation can be often used to approximate the Tweedie densities, where $b_\rho(y, \phi)$ is replaced by a simple analytic expression, which leads to a simple expression of the density [3]:

$$p(y; \mu, \phi) = (2\pi\phi y^\rho)^{-\frac{1}{2}} \exp \left(\frac{-d(y, \mu)}{2\phi} \right). \qquad (6)$$

Table 1. Exponential dispersion models with power variance functions.

Distribution	ρ	$V[\mu]$	ϕ
Gaussian	0	1	σ^2
Poisson	1	μ	1
Poisson-Gamma	$1 < \rho < 2$	μ^ρ	ϕ
Gamma(α, β)	2	μ^2	$1/\alpha$
Inverse Gaussian	3	μ^3	ϕ

In the following section, we show that (6) is the key to derive a universal denoising formula that can be used for various special cases of Tweedie distribution.

3.2. Noise2Score for Tweedie distribution

Now using the same idea of Noise2Score that relies on (3), we can obtain the following universal denoising formula from the saddle point approximation (6). All the proofs in this section are deferred to Supplementary Material.

Proposition 1 *For the given measurement model* (6)*, the Bayes optimal estimate of the unknown μ is given by*

$$\hat{\mu} = \mathbb{E}[\mu|y] = y \left(1 + (1 - \rho)\alpha(y, \rho, \phi)\right)^{\frac{1}{1-\rho}} \quad (7)$$

where

$$\alpha(y, \rho, \phi) = \phi y^{\rho-1} \left(\frac{\rho}{2y} + l'(y)\right).$$

Proposition 2 *The estimate* (7) *converges to the specific formulae in Table 2 for given parameter pairs (ρ, ϕ).*

Table 2. Special cases of Tweedie's formula for denoising.

Distribution	ρ	ϕ	$\hat{\mu}$
Gaussian	0	σ^2	$y + \sigma^2 l'(y)$
Poisson	1	ζ	$\left(y + \frac{\zeta}{2}\right)\exp(\zeta l'(y))$
Gamma(α, α)	2	$1/\alpha$	$\frac{\alpha y}{(\alpha-1) - y l'(y)}$

Note that the specific formula in Table 2 is equivalent to that of the original paper of Noise2Score [10]. This suggests that by simply estimating the parameter pair (ρ, ϕ), we can estimate both noise model and noise levels. In the following, we provide a systematic algorithm to estimate these parameters from a given noisy measurement.

3.3. Noise Model and Level Estimation

3.3.1 Noise model estimation

Here, we provide an algorithm that can be used to estimate the noise model parameter ρ. Let y_1 be the noisy measurement. Suppose we add a small amount of independent noise

to generate y_1:

$$y_2 = y_1 + \epsilon u, \quad u \sim \mathcal{N}(0, I)$$

where ϵ is a small known value. If the injected noises are sufficiently small, one could expect that their denoised images using (7) should be similar, i.e.

$$\mathbb{E}[\mu \mid y_1] \simeq \mathbb{E}[\mu \mid y_2] \quad (8)$$

Therefore, we have $\alpha(y_1, \rho, \phi) \simeq \alpha(y_2, \rho, \phi)$, which is equivalent to:

$$\phi y_1^{\rho-1} \left(\frac{\rho}{2y_1} + l'(y_1)\right) \simeq \phi y_2^{\rho-1} \left(\frac{\rho}{2y_2} + l'(y_2)\right) \quad (9)$$

Now, the key observation is that the noise level parameter ϕ can be canceled from (9) when the equality holds. Therefore, (9) provides a closed form formula for the noise model parameter ρ as stated in the following proposition.

Proposition 3 *Suppose that equality holds in* (9) *for $y_1 \neq y_2$. Then, the distribution parameter ρ is given by*

$$\hat{\rho} = \frac{-a(b-2) \pm \sqrt{(a(b-2)^2 - 4a(-2ab+w)}}{2a} \quad (10)$$

where $a = \log(\frac{y_2}{y_1}), b = 2y_1 l'(y_1)$ and $w = 2y_2 l'(y_2) - 2y_1 l'(y_1)$.

Proposition 3 implies that we can estimate the unknown noise model by solving the quadratic equation. Among two solutions for ρ, we empirically determine $\hat{\rho}$ by taking the maximum of two values for the correct estimate of ρ. Since we assume that $\rho \in (0, 2)$, we set the final $\hat{\rho}$ to $\max(0, \hat{\rho})$.

To determine whether underlying noises are either from Gaussian, Poisson, and Gamma, we set the empirical rule to estimate the noise model from $\hat{\rho}$:

$$\begin{cases} \text{Gaussian,} & \text{if } 0 \leq \hat{\rho} < 0.9, \\ \text{Poisson,} & \text{if } 0.9 \leq \hat{\rho} < 1.9, \\ \text{Gamma,} & \text{if } 1.9 \leq \hat{\rho} < 2.9. \end{cases} \quad (11)$$

Specifically, if the estimated model parameter $\hat{\rho}$ is included in a specific range, we conclude that the noisy images belong to corresponding the noise distribution. The detail of implementation are described in Supplementary Material.

3.3.2 Noise level estimation

If the noise model parameter ρ is known or correctly estimated, the universal denoising formula in (7) still require the knowledge of noise level ϕ.

Although the original Noise2Score could also estimates noise level using image quality penalty metric such as total variation (TV), our method can estimate the unknown noise

parameter much more efficiently. Again, the basic intuition is that if the injected noises are sufficiently small, one could expect that their denoised images using (7) should be similar. Proposition 4 provides the specific formula based on the assumption.

Proposition 4 *Assume that the noise model parameter ρ is known. Suppose, furthermore, that the equality holds in (8) for $y_1 \neq y_2$. Then, the noise level parameter of several noise distribution is given by*

$$
\begin{cases}
\hat{\sigma}^2 = \left(\frac{-\epsilon u}{l'(y_2) - l'(y_1)} \right), & Gaussian, \\
\hat{\zeta} = \left(-y_1 + \sqrt{y_1^2 - 2c} \right), & Poisson, \\
\hat{\alpha} = \hat{\beta} = \hat{k} = \left(1 + \frac{l'(y_2) - l'(y_1)}{\frac{1}{y_2} - \frac{1}{y_1}} \right), & Gamma,
\end{cases}
\tag{12}
$$

where $c = \epsilon u/(l'(y_2) - l'(y_1))$.

For the case of Poisson noise, the solution is derived from the quadratic equation for ζ. Due to $\zeta > 0$, we only calculate the solution with a positive sign. All these formulae are applied for each pixel level, so we should use the median for the pixel level estimate.

In contrast to Noise2Score which requires multiple inferences for the noise level estimation, our proposed method requires only one more inference step to find unknown noise level parameter. Thus, the proposed method has advantages in terms of speed and simplicity.

4. Experimental Results

4.1. Implementation Details

Training Details In order to fairly compare the proposed method with other comparison methods, a modified version of U-Net generator [12] is used for all methods. The minibatch size was adopted as 16 and the total epoch was set to 100 for training. The Adam optimizer [11] was used to train the neural network with an initial learning rate 2×10^{-4}. The learning rate was decayed to 2×10^{-5} after 50 epochs. To train the score model in this work, we employed the AR-DAE in Noise2Score method [10] and adopt the exponential moving average method with decay rate 0.999. For the annealing sigma σ_a which is required to learn the score function of noisy data, we generate geometric sequences which have the value of the perturbed noise level from σ_a^{max} to σ_a^{min} [22]. Further details of implementation are described in Supplementary Material. Our method was implemented in PyTorch [19] with NVIDIA GeForce GTX 2080-Ti.

Datasets for Synthetic Experiments We used DIV2K [23] and CBSD400 [16] dataset as training data set. We sampled 220,000 cropped patches of 128×128 size to train the network. For the data augmentation, we

used the random horizontal, vertical flip, and flop methods. We generated the synthetic noise images for blind noise cases where noise levels are randomly sampled from the particular range for each noise distribution. In the case of Gaussian noise, noise levels varied with $\sigma \in [5,55]$. For Poisson noise, noise levels varied with $\zeta \in [0.1, 0.005]$. For Gamma noise, noise levels are in $k \in [40, 120]$. To evaluate the proposed method for blind synthetic noise experiment, the test set was adopted for Kodak, CBSD68 [16] dataset.

4.2. Results of Synthetic Experiments.

Gaussian noise To evaluate our method on various noise distribution, we adopted the nine comparison methods such as BM3D [2], N2V [12], N2S [1], Nei2Nei [6], Laine19 [13], N2Score [10], N2N [14], and supervised learning approaches as shown in Table 3. Comparison methods are varied depending on the noise distribution. The supervised learning and N2N using multiple pair images perform best, but are not practical. In the case of additive Gaussian noise, our method not only outperforms the other self-supervised learning approaches for all of dataset but also provides comparable results to the supervised learning approaches. Thanks to the improved score function estimation described in Supplementary Material, our method provides even better performance of N2Score. The qualitative comparison in Figs. 3 and 4 show that the proposed method provides the best reconstruction results.

Poisson noise For the Poisson noise case, BM3D was replaced with Anscombe transformation (BM3D+VST) as indicated in Table 3. Our proposed method provides significant gain in performance compared to other algorithms, although Laine19-pme, taking advantage of the noise model with known prior, shows the best performance for the Poisson case. Note that our proposed method assumes that the noise model is unknown and estimates the noise statistics. Despite the unknown noise model, the results of proposed method are still comparable. We found that our method significantly improved the results of Noise2Score in Poisson case when the noise level parameter is unknown. It confirmed that the proposed noise level estimation method is more effective than quality penalty metrics in Noise2Score. The visual comparison results in Figs. 3 and 4 show that our method delivers much visually pleasing results compared to other self-supervised methods.

Gamma noise As the extension of BM3D and Laine19 for Gamma noises are not available, we only used six comparison methods to evaluate the proposed method, as indicated in Table 3. In the case of Gamma noise case, we set $\alpha = \beta = k$. Again, the proposed method yields the best results against self-supervised learning approaches and significantly improves the performance of Noise2Score with a margin of $\sim$1dB. The qualitative comparison in Figs. 3
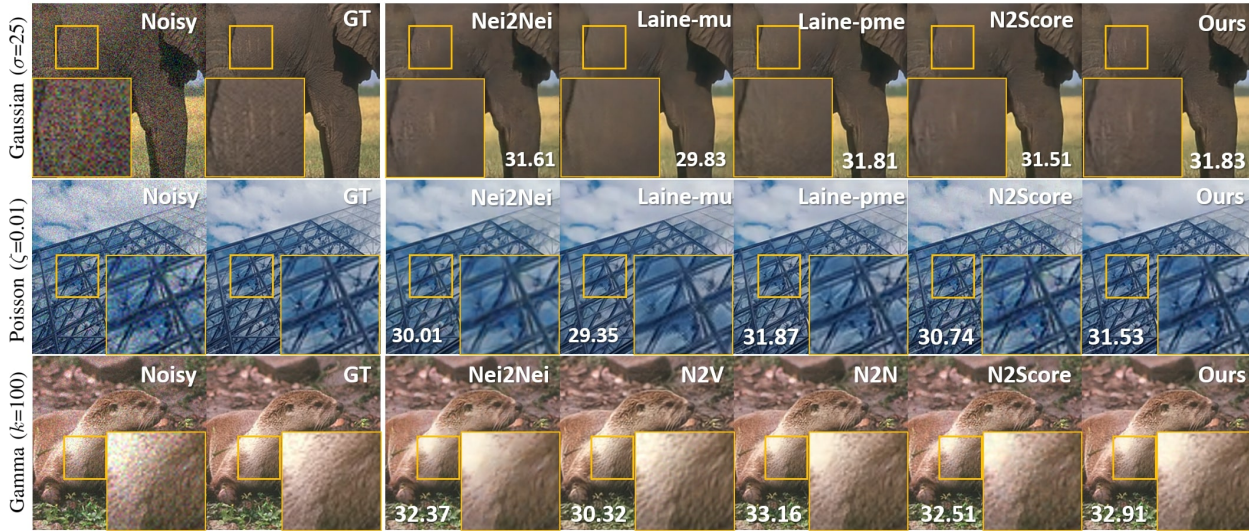
Figure 3. Visual comparison of our method against other competing methods in CBSD68 dataset. Top : Gaussian noise with $\sigma = 25$. Middle: Poisson noise with $\zeta = 0.01$. Bottom: Gamma noise with $k = 100$. The number at the lower corner of images indicates the PSNR value in dB.
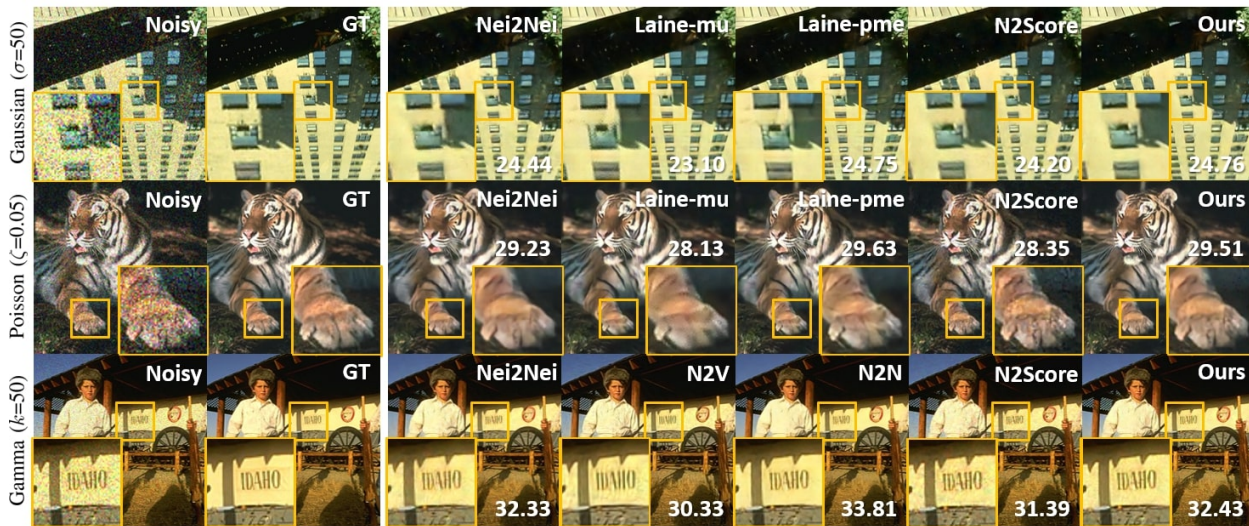

Figure 4. Visual comparison of our method against other competing methods in CBSD68 dataset. Top : Gaussian noise with $\sigma = 50$. Middle: Poisson noise with $\zeta = 0.05$. Bottom: Gamma noise with $k = 50$. The number at the lower corner of the image indicates the PSNR value in dB.

and 4 confirm that our method provides competitive visual quality among self-supervised deep denoisers.

4.3. Real Image Noise Removal

To verify that the algorithm can be applied to a real dataset, a real noise removal experiment was performed using the AAPM CT dataset [17], which contains 4358 images in 512×512 resolution from 10 different patients at low-dose level and high-dose X-ray levels. The input images are

adopted as quarter dose images and the target images are set to full dose images. We randomly select 3937 images as a train set, and the remaining 421 images are set as a test set. The noise distribution in X-ray photon measurement and in the sinogram are often modeled as Poisson noise and Gaussian noise, but real noise in reconstructed images is spatially correlated and more complicated, so that it leads to difficulties in modeling the specific noise distributions.

To deal with blind image noise removal in the low dose

Table 3. Quantitative comparison of denoising results by various methods in terms of PSNR(dB) when the noise parameters are unknown (N2V: Noise2Void, N2S: Noise2Self, Nei2Nei: Neighbor2Neighbor, N2Score: Noise2Score, N2N: Noise2Noise, SL: supervised learning, Anscombe: Anscombe method for BM3D).

| Noise type		Conventional	Self-Supervised							Supervised	
Gaussian	Dataset	BM3D	N2V	N2S	Nei2Nei	Laine19-mu	Lain19-pme	N2Score	Ours	N2N	SL
$\sigma = 25$	CBSD68	28.61	29.22	29.71	30.56	28.61	30.88	30.75	**30.89**	30.92	30.92
	Kodak	29.94	30.02	30.81	31.55	30.19	31.92	31. 78	**31.95**	31.96	31.96
$\sigma = 50$	CBSD68	26.71	25.13	27.14	27.32	26.42	**27.65**	27.32	27.56	27.73	27.73
	Kodak	27.02	25.75	28.21	28.28	27.78	28.63	28.23	**28.64**	28.70	28.71
Poisson	Dataset	Anscombe	N2V	N2S	Nei2Nei	Laine19-mu	Lain19-pme	N2Score	Ours	N2N	SL
$\zeta = 0.01$	CBSD68	30.68	31.02	30.74	31.64	30.61	**32.73**	31.87	32.53	32.94	32.95
	Kodak	31.93	31.98	31.92	32.61	31.67	**33.51**	32.96	33.41	33.86	33.87
$\zeta = 0.05$	CBSD68	26.93	28.12	28.33	28.46	27.53	**28.83**	28.27	28.73	29.07	29.07
	Kodak	28.27	29.32	29.49	29.54	28.68	**30.17**	28.98	29.72	30.23	30.25
Gamma	Dataset		N2V	N2S	Nei2Nei			N2Score	Ours	N2N	SL
$k = 100$	CBSD68	-	31.83	31.71	34.21	-	-	33.82	**34.52**	35.33	35.33
	Kodak	-	31.66	32.83	35.10	-	-	34.23	**35.42**	36.16	36.16
$k = 50$	CBSD68	-	30.51	30.63	32.11	-	-	31.32	**32.43**	33.41	33.41
	Kodak	-	31.38	31.71	32.38	-	-	31.81	**32.81**	34.39	34.40
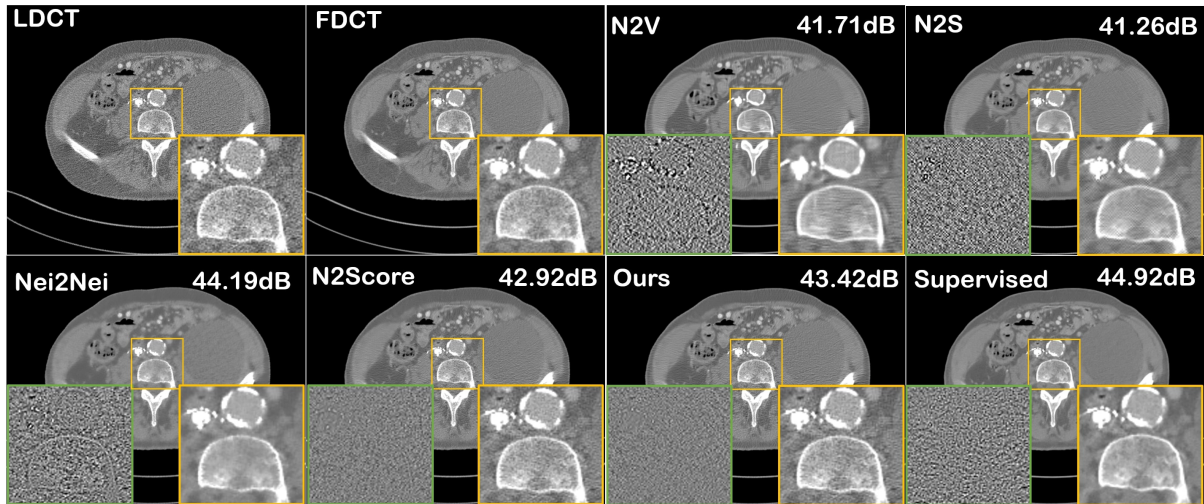
Figure 5. Denoising results of AAPM data using various methods. The yellow box and green box show the enlarged view of image and difference image between network input and output, respectively. The intensity window of CT image is (-500,500)[HU] and the intensity window of difference is (-200,200) [HU].

CT images, we carried out experiment with the proposed method using the procedure in Fig. 2. The other experiment settings are identical to synthetic experiments. Based on the model estimation rule in (11), we observed that the Low-dose CT images can be interpreted as corrupted by Gaussian noise ($\hat{\rho} \simeq 0$). To evaluate the proposed method, we compare with other methods such as N2S [1], N2V [12], Nei2Nei [6], N2Score [10], and supervised learning approach. To implement the original Noise2Score, we assume that the noise distribution is additive Gaussian noise. The quantitative results suggest that the proposed method yields

the competitive results to other self-supervised learning approaches. However, the visual quality is also an important criterion to evaluate the results in the medical imaging. In Fig. 5, the qualitative comparison shows that other self-supervised approaches provide over-smooth denoised images as shown in Table 4. In particular, the difference images of these methods show the structure of CT images. However, our method yields visually similar results compared to full-dose CT image and also provides only noises in the difference image between the network input and output. Therefore, our method also has the advantage that it
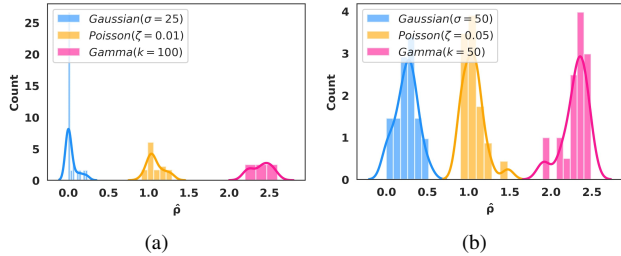
Figure 6. The histogram of estimated noise model parameter $\hat{\rho}$ in Kodak dataset. (a) For the case of σ =25, ζ = 0.01, k = 100. (b) For the case of σ=50, ζ = 0.05, k = 50.

can be applied to real environment dataset.

Table 4. Quantitative comparison using various methods in terms of PNSR(dB) on the AAPM CT dataset. (N2V: Noise2Void, N2S: Noise2Self, Nei2Nei: Neighbor2Neighbor, N2Score: Noise2Score SL: supervised learning).

Method	N2S	N2V	Nei2Nei	N2Score	Ours	SL
PSNR	35.82	35.93	37.83	36.63	36.92	38.53

5. Ablation Study

5.1. Results on Noise Model Estimation

Here, we validate the proposed the noise model estimation by (11). Figure 6 shows the histogram of estimated noise model parameter $\hat{\rho}$ in the Kodak dataset. We have calculated $\hat{\rho}$ for each noise distribution and plot distributions in each figure. According to Tweedie's distribution, we expect that the parameters of noise model can be differentiated for each noise distribution such as Gaussian ($\hat{\rho} \simeq 0$), Poisson ($\hat{\rho} \simeq 1$), and Gamma ($\hat{\rho} \simeq 2$). From the figure, we can observe that the noise model parameters $\hat{\rho}$ are distinctly distributed in the case of low noise levels. Furthermore, in the case of high noise levels, the distributions can also be distinguished for each noise distribution. It implies that it was possible to successfully estimate the noise model with the proposed method, so that it leads to the use of the exact Tweedie's formula for blind image denoising.

Table 5. Reconstruction results with various noise parameter estimation methods on CBSD68 dataset in terms of PSNR(dB).

Noise type	N2Score	Ours	Known value	SL
Gaussian (σ=25)	30.78	30.89	30.91	30.92
Poisson (ζ=0.01)	31.89	32.53	32.63	32.95
Gamma (k=100)	33.92	34.52	34.53	35.33
Inference speed	3.1s	0.1s	-	-

5.2. Ablation Study on Noise Level Estimation

We analyze the effect of the proposed noise level estimation by comparing Noise2Score with quality penalty metric [10] and others. To fairly evaluate the proposed method,

we used an identical weights of the score model for each case. We carried out an ablation study by fixing the noise model estimation procedure and by only varying the noise level estimation as shown Table 5. Here, "known value" refers to the results assuming the known ground truth noise level parameter. "SL" denotes that the results of supervised learning. We found that quality penalty metric yield comparable performance to "known value" in the case of the Gaussian, but the performance with the quality penalty metric decreased for the case of Poisson and Gamma noises. However, the proposed methods outperform the quality penalty metric approach and yields results comparable to "known value" in all cases. Furthermore, we compare the inference speed of the estimation stage. The Noise2Score took about $\times$ 30 times more to estimate noise level parameters compared to our method.

6. Conclusion

In this article, we provided a novel self-supervised blind image denoising framework that does not require clean data and prior knowledge of noise models and levels. Our innovation came from the saddle point approximation of Tweedie distributions, which cover a wide range of exponential family distributions. By taking advantage of this property, we provided a universal denoising formula that can be used for various distributions in real life. Furthermore, we proposed a novel algorithm that can estimate the noise model and noise level parameter in a unified framework. Finally, we validated the proposed method using benchmark and real CT image data sets, and confirmed that the method outperforms the existing state-of-the-art self-supervised learning methods.

Limitation

While we provide a unified approach for noise distribution adaptive self-supervised image denoising, the method is not free of limitations. As the formulae for estimating of noise model and levels were derived from several approximations, it could fail in a real-world dataset. Furthermore, the noise model may not be described by Tweedie distribution in real environments.

Acknowledgement

References

[1] Joshua Batson and Loic Royer. Noise2Self: Blind denoising by self-supervision. In *International Conference on Machine Learning*, pages 524–533. PMLR, 2019. 2, 5, 7

[2] Kostadin Dabov, Alessandro Foi, Vladimir Katkovnik, and Karen Egiazarian. Image denoising with block-matching and 3D filtering. In *Image Processing: Algorithms and Systems, Neural Networks, and Machine Learning*, volume 6064, page 606414. International Society for Optics and Photonics, 2006. 1, 5

[3] Peter K Dunn and Gordon K Smyth. Tweedie family densities: methods of evaluation. In *Proceedings of the 16th international workshop on statistical modelling, Odense, Denmark*, pages 2–6. Citeseer, 2001. 3

[4] Bradley Efron. Tweedie's formula and selection bias. *Journal of the American Statistical Association*, 106(496):1602–1614, 2011. 3

[5] Shuhang Gu, Lei Zhang, Wangmeng Zuo, and Xiangchu Feng. Weighted nuclear norm minimization with application to image denoising. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2862–2869, 2014. 1

[6] Tao Huang, Songjiang Li, Xu Jia, Huchuan Lu, and Jianzhuang Liu. Neighbor2Neighbor: Self-Supervised Denoising from Single Noisy Images. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 14781–14790, 2021. 2, 5, 7

[7] Aapo Hyvärinen and Peter Dayan. Estimation of nonnormalized statistical models by score matching. *Journal of Machine Learning Research*, 6(4), 2005. 2

[8] Bent Jorgensen. *The theory of dispersion models*. CRC Press, 1997. 3

[9] Kwanyoung Kim, Shakarim Soltanayev, and Se Young Chun. Unsupervised Training of Denoisers for Low-Dose CT Reconstruction Without Full-Dose Ground Truth. *IEEE Journal of Selected Topics in Signal Processing*, 14(6):1112–1125, 2020. 2, 3

[10] Kwanyoung Kim and Jong Chul Ye. Noise2Score: Tweedie's Approach to Self-Supervised Image Denoising without Clean Images. *arXiv preprint arXiv:2106.07009*, 2021. 2, 3, 4, 5, 7, 8

[11] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014. 5

[12] Alexander Krull, Tim-Oliver Buchholz, and Florian Jug. Noise2Void-learning denoising from single noisy images. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2129–2137, 2019. 2, 5, 7

[13] Samuli Laine, Tero Karras, Jaakko Lehtinen, and Timo Aila. High-quality self-supervised deep image denoising. *arXiv preprint arXiv:1901.10277*, 2019. 2, 5

[14] Jaakko Lehtinen, Jacob Munkberg, Jon Hasselgren, Samuli Laine, Tero Karras, Miika Aittala, and Timo Aila. Noise2Noise: Learning image restoration without clean data. *arXiv preprint arXiv:1803.04189*, 2018. 2, 5

[15] Jae Hyun Lim, Aaron Courville, Christopher Pal, and Chin-Wei Huang. AR-DAE: Towards Unbiased Neural Entropy Gradient Estimation. In *International Conference on Machine Learning*, pages 6061–6071. PMLR, 2020. 3

[16] 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 *Proceedings Eighth IEEE International Conference on Computer Vision. ICCV 2001*, volume 2, pages 416–423. IEEE, 2001. 5

[17] Cynthia H McCollough, Adam C Bartley, Rickey E Carter, Baiyu Chen, Tammy A Drees, Phillip Edwards, David R Holmes III, Alice E Huang, Farhana Khan, Shuai Leng, et al. Low-dose CT for the detection and classification of metastatic liver lesions: results of the 2016 low dose CT grand challenge. *Medical physics*, 44(10):e339–e352, 2017. 6

[18] Nick Moran, Dan Schmidt, Yu Zhong, and Patrick Coady. Noisier2noise: Learning to denoise from unpaired noisy data. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 12064–12072, 2020. 2

[19] 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. 2017. 5

[20] Yuhui Quan, Mingqin Chen, Tongyao Pang, and Hui Ji. Self2self with dropout: Learning self-supervised denoising from single image. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1890–1898, 2020. 2

[21] Shakarim Soltanayev and Se Young Chun. Training and Refining Deep Learning Based Denoisers without Ground Truth Data. *arXiv preprint arXiv:1803.01314*, 2018. 2, 3

[22] Yang Song and Stefano Ermon. Improved techniques for training score-based generative models. *arXiv preprint arXiv:2006.09011*, 2020. 5

[23] Radu Timofte, Shuhang Gu, Jiqing Wu, Luc Van Gool, Lei Zhang, Ming-Hsuan Yang, Muhammad Haris, et al. NTIRE 2018 challenge on Single Image Super-Resolution: Methods and Results. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, June 2018. 5

[24] Pascal Vincent, Hugo Larochelle, Isabelle Lajoie, Yoshua Bengio, Pierre-Antoine Manzagol, and Léon Bottou. Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion. *Journal of machine learning research*, 11(12), 2010. 3

[25] Yaochen Xie, Zhengyang Wang, and Shuiwang Ji. Noise2Same: Optimizing A Self-Supervised Bound for Image Denoising. *arXiv preprint arXiv:2010.11971*, 2020. 2