Theoretical analysis on Noise2Noise using Stein’s Unbiased Risk Estimator for Gaussian denoising: Towards unsupervised training with clipped noisy images

Magauiya Zhussip 1 Shakarim Soltanayev 1 Se Young Chun 1

Abstract

Recently, Noise2Noise has been proposed for unsupervised training of deep neural networks in image restoration problems including denoising Gaussian noise. However, it does not work well for truncated noise with non-zero mean. Here, we perform theoretical analysis on Noise2Noise for the limited case of Gaussian noise removal using Stein’s Unbiased Risk Estimator (SURE). We extend SURE to deal with a pair of noise realizations to directly compare with Noise2Noise. Then, we show that Noise2Noise with Gaussian noise is a special case of our newly extended SURE with a pair of uncorrelated noise realizations. Lastly, we propose a compensation method for clipped Gaussian noise to approximately follow Normal distribution and show how this compensation method can be used for SURE based unsupervised denoiser training. We also show that our theoretical analysis provides insights on how to use Noise2Noise for clipped Gaussian noise.

1. Introduction

Deep learning based approaches have been quite successful in high-level computer vision tasks such as image classification (Krizhevsky et al., 2012; He et al., 2016), object detection (Ren et al., 2015; Redmon & Farhadi, 2017), and semantic segmentation (Long et al., 2015). These data-driven methods that are trained with images and ground truth labels outperform conventional methods by large margins. There have been several works on using deep neural networks (DNNs) for low-level computer vision problems such as single image super resolution (Kim et al., 2016; Ledig et al., 2017), image inpainting (Xie et al., 2012), image restoration (Mao et al., 2016) and image denoising (Burger et al., 2012; Zhang et al., 2017). These methods are using ground truth images to train DNNs and most of them outperform conventional approaches. However, image denoising is somewhat challenging for deep learning based approaches to win over conventional methods such as BM3D (Dabov et al., 2007), especially for (clipped) noise in real photographs (Plötz & Roth, 2017). It was recent that DNN based denoisers outperformed BM3D for synthetic, unclipped Gaussian noise (Zhang et al., 2017).

However, it is challenging to obtain a massive amount of clean, noiseless ground truth images for training DNN based denoisers in some application areas, such as hyperspectral remote sensing and medical imaging, where the acquisition of noiseless ground truth data is expensive, or sometimes even infeasible. Thus, it is often desirable to train DNN based denoisers without clean ground truth image while the trained DNN outperforms state-of-the-art conventional methods. To address this issue, there have recently been several works on unsupervised learning for deep learning based denoisers or unsupervised approach for DNN based image denoising. Deep image prior does not require training process, but performs the mean-squared error (MSE) minimization between the output of a DNN and the given noisy image to denoise (Ulyanov et al., 2018).

Another approach is to train denoising DNNs only with noisy images. Noise2Noise was proposed to train image restoration DNNs with the set of a pair of noisy images for a clean image in an unsupervised way (Lehtinen et al., 2018) for a wide variety of noises and MR reconstruction. Stein’s unbiased risk estimator (SURE) based training method for Gaussian denoisers was also developed (Soltanayev & Chun, 2018) to train DNNs with the set of a single noise realization for an image in an unsupervised way. The SURE based method is to use a single noise realization per image for training, but Noise2Noise does use a pair of noise realizations per image. However, Noise2Noise has much wider applications in image denoising and restoration than SURE method does since the latter is limited to Gaussian noise. In summary, both Noise2Noise and SURE method outperform conventional methods such as BM3D for zero-mean noise and synthetic Gaussian noise, respectively. They often yield comparable performances to DNN based denoisers that are trained with ground truth. However, none of them is working for more realistic noise such as truncated Gaussian with...
non-zero mean (Plötz & Roth, 2017; Lehtinen et al., 2018).

In this article, we perform theoretical analysis to deepen our understandings on Noise2Noise (Lehtinen et al., 2018) and SURE based method (Soltanayev & Chun, 2018) for limited cases of Gaussian denoising. Since Noise2Noise takes a pair of noisy images and SURE takes a single noisy image, we propose to modify SURE so that it can deal with a pair of noisy images (we will denote it by eSURE or extended SURE). Then, we will prove that there is a clear theoretical link between Noise2Noise and extended SURE. In fact, we will show that Noise2Noise is a special case of extended SURE for (unclipped) uncorrelated Gaussian noise in theory. Lastly, we propose a compensation method for clipped Gaussian noise to approximately follow Normal distribution and show how this compensation method can be used for SURE based unsupervised denoiser training. We will show that our theoretical analysis provides insights on how to use Noise2Noise for clipped Gaussian noise. Simulations are performed with MNIST, BSD68 and Set12 datasets to show that our developed theories are consistent with simulations.

2. Background

2.1. Stein’s unbiased risk estimator (SURE)

Typically, Gaussian contaminated signal (or image) is modeled as a linear equation:

\[ y = x + n \]  

(1)

where \( x \in \mathbb{R}^N \) is an unknown signal, \( y \in \mathbb{R}^N \) is a known measurement, \( n \in \mathbb{R}^N \) is an i.i.d. Gaussian noise such that \( n \sim N(0, \sigma^2 I) \), and \( I \) is an identity matrix. We denote \( n \sim N(0, \sigma^2 I) \) as \( n \sim \mathcal{N}_0, \sigma^2 \).

In general, given an estimator \( h(y) \) of \( x \), the SURE has the following form:

\[
\eta(h(y)) = \frac{\|y - h(y)\|^2}{N} - \sigma^2 + \frac{2\sigma^2}{N} \sum_{i=1}^{N} \frac{\partial h_i(y)}{\partial y_i} \]  

(2)

Assuming \( x \) to be deterministic signal (or image), we can estimate the expectation of an MSE by utilizing the following theorem:

**Theorem 1.** (Stein, 1981; Blu & Luisier, 2007) The random variable \( \eta(h(y)) \) is an unbiased estimator of

\[
\text{MSE}(h(y)) = \frac{1}{N} \|x - h(y)\|^2
\]

or

\[
\mathbb{E}_{n \sim \mathcal{N}_0, \sigma^2} \left\{ \frac{\|x - h(y)\|^2}{N} \right\} = \mathbb{E}_{n \sim \mathcal{N}_0, \sigma^2} \{ \eta(h(y)) \} \]  

(3)

where \( \mathbb{E}_{n \sim \mathcal{N}_0, \sigma^2} \{ \cdot \} \) is the expectation operator in terms of the random vector \( n \).

Although (2) seems appealing in terms of optimizing parameters of an estimator \( h(y) \), analytical solution for the last term (divergence) is often available only for some limited, special cases such as non-local mean or linear filters (Van De Ville & Kocher, 2009; 2011). Thus, one needs to find at least an approximate solution for the general case. It may also be possible to calculate the divergence term using deep learning software packages such as Tensorflow. However, note that one need to calculate the derivative of the divergence term in order to use SURE for training deep learning based denoisers.

2.2. Monte-Carlo SURE (MC-SURE)

A fast Monte-Carlo approximation of the divergence term has been developed in (Ramani et al., 2008). The introduced method yielded accurate unbiased estimate of MSE for many denoising methods \( h(y) \).

**Theorem 2.** (Ramani et al., 2008) Let \( \tilde{n} \sim \mathcal{N}_{0,1} \in \mathbb{R}^N \) be independent of \( n \) or \( y \). Then,

\[
\sum_{i=1}^{K} \frac{\partial h_i(y)}{\partial y_i} = \lim_{\epsilon \to 0} \mathbb{E}_{\tilde{n}} \left\{ \tilde{n}^T \left( \frac{h(y + \epsilon \tilde{n}) - h(y)}{\epsilon} \right) \right\}
\]

(4)

provided that \( h(y) \) admits a well-defined second-order Taylor expansion. If not, this is still valid in the weak sense provided that \( h(y) \) is tempered.

Consequently, by applying Theorem 2 to the divergence term in (2), the divergence approximation of the denoiser \( h(y) \) will be:

\[
\frac{1}{N} \sum_{i=1}^{N} \frac{\partial h_i(y)}{\partial y_i} \approx \frac{1}{\epsilon N} \tilde{n}^T (h(y + \epsilon \tilde{n}) - h(y)),
\]

(5)

where \( \tilde{n}^T \) is a transposed \( i.i.d \) Gaussian vector \( \tilde{n} \sim \mathcal{N}_{0,1} \) and \( \epsilon \) is a fixed small positive value.

2.3. SURE based deep denoiser training

Recently, SURE was used as the metric to approximate and minimize the MSE between the output of the DNNs and the ground truth (Soltanayev & Chun, 2018). Specifically, MC-SURE allows DNN to learn large-scale weights in a DNN by minimizing MC-SURE without noisless ground truth images. The equation (2) with (4) was reformulated for a DNN as follows:

\[
\eta(h_\theta(y)) = \frac{1}{M} \sum_{j=1}^{M} \left\{ \|y^{(j)} - h_\theta(y^{(j)})\|^2 - N\sigma^2 \right\}
\]

\[
+ 2\sigma^2 \left( \tilde{n}^{(j)} \right)^T \left( h_\theta(y^{(j)}) + \epsilon \tilde{n}^{(j)} - h_\theta(y^{(j)}) \right),
\]

(6)
where \( \theta \) is a set of denoiser parameters, \( M \) is a size of mini-batch, \( \epsilon \) is a small fixed positive constant, and \( \tilde{n}^{(j)} \) is a single realization from the standard normal distribution for each training data \( j \). This approach was demonstrated to yield state-of-the-art performance in denoising task with ideal Gaussian noise, having almost similar qualitative and quantitative results compared to the trained DNNs with MSE. However, MC-SURE based DNN training method is limited to \( i.i.d. \) Gaussian noise contaminated images, while in practice noisy images are clipped due to the hardware limitations or saturation.

### 3. Methods

In this section, we extend SURE and MC-SURE to deal with a pair of noisy images instead of a single noisy image and use it for training deep learning based denoisers under the assumption of \( i.i.d. \) Gaussian noise model. Moreover, we prove that Noise2Noise is a special case of our proposed extended SURE for Gaussian denoising problems. Finally, it will be demonstrated that our method can further be extended to the case with clipped noisy images, which follow a so-called truncated Gaussian noise model.

#### 3.1. Extended SURE and MC-SURE

The proposed SURE with a pair of noisy images is formulated in the following way:

**Theorem 3.** Let \( y_1 \sim N(x, \sigma^2_y I) \) and \( y_2 \sim N(0, \sigma^2_y I) \) be a pair of noisy images of the realizations.

Then, the random variable \( \gamma(h_\theta(y_2), y_1) \) is an unbiased estimator of MSE:

\[
E_{y_2} \left\{ \frac{1}{N} \| x - h_\theta(y_2) \|^2 \right\} = E_{y_2} \{ \gamma(h_\theta(y_2), y_1) \}
\]

where

\[
\gamma(h_\theta(y_2), y_1) = \frac{1}{N} \| y_1 - h_\theta(y_2) \|^2 - \sigma^2_y
\]

**Proof.**

\[
E_{y_2} \left\{ \frac{1}{N} \| x - h_\theta(y_2) \|^2 \right\}
\]

\[
= E_{y_1, y_2} \left\{ \frac{1}{N} \| y_1 - h_\theta(y_2) \|^2 \right\} - \sigma^2_y
\]

\[
+ \frac{2}{N} E_{y_1, y_2} \left\{ (y_1 - x)^T h_\theta(y_2) \right\}
\]

\[
E_{y_2} \left\{ \sum_{i=1}^{N} \frac{\partial h_i(y_2)}{\partial y_2} \right\}
\]

\[
= \frac{1}{\sigma^2_y} E_{y_1, y_2} \left\{ (y_1 - x)^T h_\theta(y_2) \right\}
\]

See the supplementary material for more details. \( \square \)

Theorem 3 is developed for the general case and one can train deep network denoisers using the following corollary:

**Corollary.** Given a noisy realization pairs of a clean image \((y_3, y_4)\) from the same distribution \( N(x, \sigma^2_x I) \), we calculate less noisy image \( w = \frac{1}{2}(y_3 + y_4) \sim N(x, \frac{1}{2}\sigma^2_x I) \). Then, we add \( i.i.d. \) Gaussian noise \( z \sim N(0, \frac{1}{2}\sigma^2_z I) \) to \( w \), so that \( v = (w + z) \sim N(x, \sigma^2_z I) \). Finally, by applying Theorem 3 and replacing divergence term with its Monte-Carlo approximation \((5)\), one can minimize extended MC-SURE with respect to \( \theta \):

\[
\gamma(h_\theta(v), w) = \frac{1}{N} \| w - h_\theta(v) \|^2
\]

\[
- \frac{1}{2} \sigma^2_y + \frac{\sigma^2_z}{\epsilon N} (\tilde{n}^T (h_\theta(v + \epsilon \tilde{n}) - h_\theta(v))
\]

Note that \( v \) has the same distribution as \( y_3 \) and \( y_4 \). Thus, instead of generating \( v \), we can simply replace it with one of the realizations.
For a training dataset of noisy \( M \) pairs \( \{y^{(1)}, \ldots, y^{(M)}\} \), we generate \( \{w^{(j)}, v^{(j)}\}, j \in [1, M] \) to train deep learning based denoisers with proposed extended MC-SURE method. In simulation part, we demonstrate that our method better approximates the MSE and hence shows better performance compared to original MC-SURE based training approach for natural image denoising.

### 3.2. Link between extended SURE and Noise2Noise

The extended SURE framework that we proposed can be applied to a pair of uncorrelated Gaussian noisy images \( (y \sim \mathcal{N}(x, \sigma_y^2)) \) and \( z \sim \mathcal{N}(x, \sigma_z^2) \). In that case, the divergence term vanishes leaving us following expression:

**Theorem 4.**

\[
E_{x,y} \{\gamma(h_\theta(y), z)\} = E_{x,y} \left\{ \frac{||z - h_\theta(y)||^2}{N} \right\} - \sigma_z^2 \tag{10}
\]

**Proof.**

\[
E_y \left\{ \frac{1}{N}||x - h_\theta(y)||^2 \right\} = E_{y,z} \left\{ \frac{1}{N}||z - h_\theta(y)||^2 \right\} - \sigma_z^2 + 2 \frac{N}{E_{y,z} \{ (z - x)^T h_\theta(y) \}} = \frac{2}{N} E_{y,z} \{ (z - x)^T h_\theta(y) \} = \frac{2}{N} E_{y,x} \{ (z - x)^T \} E_{x,y} h_\theta(y).
\]

See supplementary material for more details. \( \square \)

From the above expression, one clearly sees that the first term corresponds to the cost function of Noise2Noise for \textit{i.i.d.} Gaussian noise denoising case (Lehtinen et al., 2018), while the second term \( \sigma_z^2 \) is a constant.

Minimization of (10) with respect to a set of denoiser parameters \( \theta \) should give us the same solution for both extended MC-SURE and Noise2Noise. Although it is not easy to notice the relation between two different approaches from the first sight, it turns out that Noise2Noise is a special case of proposed MC-SURE based training method for \textit{i.i.d.} Gaussian denoising task.

### 3.3. MC-SURE for clipped \textit{i.i.d.} Gaussian noise

In practice, the assumption that noisy observations follow the Gaussian noise model is partially incorrect, since they have been clipped by digital imaging sensors (e.g. CMOS imager). Thus, instead of unclipped noisy image \( y \), we observe \( \tilde{y} \):

\[
\tilde{y} = f(y) = \max\{\min\{y, 1\}, 0\} \tag{11}
\]

In that case, for a pair of clipped noisy images \( (\tilde{y}_1, \tilde{y}_2) \), the mean will deviate from clean image \( x \). Noise2Noise sticks to a strong assumption that even if noisy realizations are not from the same distribution, their expectations should match the target. Consequently, Noise2Noise cannot be applied to optimize a deep neural network for truncated \textit{i.i.d.} Gaussian denoising case, while proposed MC-SURE framework is able to be extended to that task.

The extended MC-SURE (9) as well as original MC-SURE (2018) can not be utilized for deep network optimization, because of \textit{i.i.d.} Gaussian distributed noise assumption. To solve this issue, we propose a method to compensate for boundary values in a \( y \) by adding \textit{i.i.d.} Gaussian noise so that:

\[
(\bar{y})_i = \begin{cases} 
(\tilde{y})_i, & \text{if } 0 < (\tilde{y})_i < 1 \\
-|n|, & \text{if } (\tilde{y})_i = 0 \\
|n| + 1, & \text{if } (\tilde{y})_i = 1,
\end{cases} \tag{12}
\]

where \( |n| \) is an absolute value of a \textit{i.i.d.} Gaussian noise with \( \mathcal{N}(0, \sigma^2) \) statistical parameters. The equation (12) assumes that true mean value of pixels at boundaries is the same as boundary values. This compensation method may work for MNIST dataset, where images mainly consist of zeros and ones.

On the other hand, for natural images, one needs to estimate true value of a pixel clipped at boundaries. Therefore, we utilize BM3D denoiser to clipped image and transform our compensation (12) to:

\[
(\bar{y})_i = \begin{cases} 
(\tilde{y})_i, & \text{if } 0 < (\tilde{y})_i < 1 \\
(\tilde{x})_i - |n|, & \text{if } (\tilde{y})_i = 0 \\
(\tilde{x})_i + |n|, & \text{if } (\tilde{y})_i = 1,
\end{cases} \tag{13}
\]

where \( \tilde{x} \) is BM3D denoised image. Worth to note that clipped pixels should be larger than boundary values, so (13) is iterated until compensated pixel \( \bar{y} \), exceeds beyond clipped values \([0-1]\).

### 3.4. Noise2Noise for clipped \textit{i.i.d.} Gaussian noise

Theorem 4 implies that Noise2Noise is valid if \( z \) contains zero-mean noise while \( y \) does not have to have since

\[
E_{y,x} \{ (z - x)^T \} E_{y,x} h_\theta(y) = 0 \tag{14}
\]

Thus, whether \( y \) contains zero-mean noise or not, Noise2Noise will work with \( z \) having zero-mean noise. Then there are two different training methods for Noise2Noise when clipped Gaussian noise is applied. Firstly, the input truncated signal to the DNN does not change, but the target will be converted into Gaussian-like noisy image and the will replace clipped noisy target image by compensated noisy image. Another method is to convert both noisy images into Gaussian-like noisy images using

Theoretical analysis on Noise2Noise using Stein’s Unbiased Risk Estimator for Gaussian denoising
our proposed compensation methods. Then, they will replace the input and the target images of Noise2Noise by converted images.

4. Experimental setup

In this section, we demonstrate the performance of proposed extended MC-SURE on MNIST and Berkeley’s BSD-400 datasets (Martin et al., 2001). As a deep denoising network, we used stacked denoising autoencoder (SDA) for MNIST, which consists of decoder and encoder with two convolutional layers and DnCNN for natural images (Vincent et al., 2010; Zhang et al., 2017). DnCNN consists of 20 layers of CNN with batch normalization followed by ReLU as a non-linear function.

The proposed method was compared with state-of-the-art denoising algorithms such as BM3D (Dabov et al., 2007), denoising networks trained on Noise2Noise (Lehtinen et al., 2018), MC-SURE (Soltanayev & Chun, 2018), all without ground truth data and MSE with ground truth data as gold standard. It is worth to note that $\epsilon$ in (2) and (9) should be carefully chosen for stable training and good performance. As also mentioned in (Deledalle et al., 2014), $\epsilon$ is directly proportional to noise standard deviation $\sigma$. Therefore, $\epsilon$ have been fine tuned, so that for our proposed MC-SURE, it was set to $\epsilon = 1.6 \times 10^{-4} \times \sigma$.

Results on MNIST dataset We trained SDA network on MNIST dataset for blind image denoising task. Set of 55,000 images were contaminated with $i.i.d.$ Gaussian noise with standard deviation $\sigma \in [0 – 55]$ to train SDA, while 5,000 images used for validation. All methods except BM3D were trained with the Adam optimization (Kingma & Ba, 2015) for 300 epochs (200 epochs with the learning rate of $10^{-3}$, 50 epochs with $10^{-4}$, and 50 epochs with $10^{-5}$). The performance of each method was evaluated on 10,000 test images.

The results reveal that our proposed extended MC-SURE method (eSURE) shows comparable results with other unsupervised methods for different noise levels (see Table 1 and Figure 1). Our training approach appears to be better than Noise2Noise for 0.03dB and slightly worse than original MC-SURE. In summary, all deep learning based methods yield results very close to each other and outperformed BM3D at least for $\sim 1$dB.

Results on natural image dataset We conducted experiments on natural image dataset to further investigate our extended MC-SURE and demonstrate the ability to deal with a pair of noise realizations per image and its superior performance over original MC-SURE. Following procedures described in DnCNN paper (Zhang et al., 2017), 128 $\times$ 2,919 patches with 50 $\times$ 50 size were extracted from 400 images of BSD-400. For Noise2Noise and proposed extended MC-SURE, two realizations of a noisy image were generated. As a testset, we utilized BSD68 dataset and 12 images from Set12 (Dabov et al., 2007).

In all cases, DnCNN denoisers were trained for blind denoising with noise level range of $\sigma \in [0 – 55]$ using Adam optimizer (Kingma & Ba, 2015). The initial learning rate was set to $10^{-3}$, which was dropped to $10^{-4}$ after 40 epochs and the network further trained for 10 epochs.

All experiments were implemented in Tensorflow framework (Abadi et al., 2016) and run on NVidia Titan X GPU. The computational cost for estimating proposed MC-SURE is the same as for original MC-SURE and thus both take approximately 26 hours to train DnCNN denoiser, while the training time for DnCNN-MSE and DnCNN-N2N was 17 hours.

The performance of our approach along with the state-of-the-art methods was tabulated in Table 2 and Table 3. The MSE trained DnCNN using a clean set of training images yielded the best performance compared to other methods for both BSD68 and Set12 testsets.

Quantitative analysis on BSD68 testset reveals that our proposed extended MC-SURE is consistently better than conventional BM3D for $\sim 0.5$dB and outperforms DnCNN-SURE for $\sim 0.15$dB in lower and higher noise cases. The performance gap between the proposed method, BM3D and DnCNN-SURE are still maintained even for Set12 testset. Also, our network training approach demonstrates very close quantitative results with Noise2Noise trained DnCNN (DnCNN-N2N) in both testsets.

In terms of visual comparison, our method effectively removed noise from an image, while preserving texture and edges. In Figure 2, conventional BM3D yielded comparable results, but for highly noisy input, the output appears blurrier. The same trend is observed for DnCNN-SURE, where details of the denoised test image from BSD68 were

| Methods | BM3D | SDA-SURE | SDA-N2N | SDA-eSURE | SDA-MSE |
|---------|------|----------|---------|-----------|---------|
| $\sigma = 25$ | 27.56 | 28.95 | 28.88 | 28.92 | 28.91 |
| $\sigma = 55$ | 21.27 | 24.48 | 24.44 | 24.47 | 24.48 |
| noisy images | 0 | 1 | 2 | 2 | $\infty$ |

Table 2. PSNR results of denoisers on BSD68 Dataset.

| Methods | BM3D | DnCNN-SURE | DnCNN-N2N | DnCNN-eSURE | DnCNN-MSE |
|---------|------|------------|-----------|-------------|-----------|
| $\sigma = 25$ | 28.56 | 28.93 | $\mathbf{29.07}$ | $\mathbf{29.07}$ | 29.20 |
| $\sigma = 50$ | 25.62 | 26.00 | 26.14 | $\mathbf{26.15}$ | 26.22 |
| noisy data | 0 | 1 | 2 | 2 | $\infty$ |
Theoretical analysis on Noise2Noise using Stein’s Unbiased Risk Estimator for Gaussian denoising

(a) Noisy input  (b) BM3D  (c) SDA-SURE  (d) SDA-N2N  (e) Ours  (f) SDA-MSE

Figure 1. Denoised MNIST testset images of SDA trained with various methods at $\sigma=55$ noise level.

(a) Ground Truth  (b) Noisy input  (c) BM3D / 28.88dB  (d) SURE / 29.30dB  (e) N2N / 29.47dB  (f) eSURE / 29.48dB

Figure 2. Denoised test image (BSD68) results of DnCNN with various methods at $\sigma=25$ noise level.

(a) Ground Truth  (b) Noisy input  (c) BM3D / 25.40 dB  (d) SURE / 25.83 dB  (e) N2N / 26.01 dB  (f) eSURE / 26.01 dB

Figure 3. Denoised test image (BSD68) results of DnCNN trained with various methods at $\sigma=50$ noise level.

vanished (see Figure 3). Also one may witness similar denoising performance for DnCNN-N2N and our method.

Both qualitative and quantitative superiority of the extended MC-SURE over original MC-SURE can be explained by its better approximation of MSE. We compared MSE estimated by extended MC-SURE and original MC-SURE while training DnCNN network. The results in Figure 4 revealed that proposed extended MC-SURE is more accurate in terms of MSE approximation.

Table 3. PSNR results of denoisers on Set12 Dataset.

| Methods      | BM3D | DnCNN-SURE | DnCNN-N2N | DnCNN-eSURE | DnCNN-MSE |
|--------------|------|------------|-----------|--------------|-----------|
| $\sigma = 25$ | 29.97| 30.04      | 30.30     | 30.31        | 30.42     |
| $\sigma = 50$ | 26.67| 26.87      | 27.07     | 27.07        | 27.16     |
| noisy data   | 0    | 1          | 2         | 2            | $\infty$  |

Figure 4. The comparison between actual MSE and its estimations: proposed extended MC-SURE and MC-SURE on BSD400 dataset during DnCNN training time.
Theoretical analysis on Noise2Noise using Stein’s Unbiased Risk Estimator for Gaussian denoising

**Clipped noisy image recovery** Denoising images with clipped noise becomes a more challenging task in case if ground-truth images are not available. In this section, we implement the original MC-SURE method describe to train SDA on clipped noisy MNIST dataset. Training images were firstly corrupted with i.i.d. Gaussian noise with $\sigma = 25$ and then the output was clipped between $[0 - 1]$. The SDA network was trained for 100 epochs with a learning rate of $10^{-3}$ and then 50 epochs with $10^{-4}$ using Adam optimizer.

We compared our SDA-eSURE (proposed) with BM3D, SDA-N2N, and SDA-SURE trained/tested on unclipped noisy images. From Table 4, we notice that our SDA-eSURE with compensation was able to achieve original SDA-SURE and outperforms other methods by orders of magnitude. Also, our method produced visually appealing results on clipped noisy MNIST testset images (see Figure 5). Note also that SDA-N2N-2 with 1 unclipped and 1 clipped noisy images also yielded comparable performance to the oracle SDA-SURE with 1 unclipped noisy image. This implies that our theoretical interpretation for Noise2Noise using SURE framework was valid and now it is possible to train the DNN with the direct input of clipped noisy images after training, while our SDA-eSURE still has to use clip compensation method during and after training.

In case of natural images, we utilized (13) to generate noise compensated BSD400 images and patchify them to $50 \times 50$ sized 180,000 patches. Training procedure was the same as for DnCNN trained unclipped noisy images.

For quantitative comparison, we included results of BM3D (Dabov et al., 2007), BM3D with homomorphic transformation (Foi, 2009) (denoted as BM3D-H), and DnCNN-SURE trained on unclipped noisy images. Table 5 demonstrates that our proposed method for clipped noisy images was able to outperform conventional methods and achieve about the same PSNR as DnCNN-SURE (no clip) that is the gold standard result. Thus, with a single realization of clipped noisy images, one can train deep learning denoiser with MC-SURE by using our compensation method.

### Table 4. PSNR results of denoisers on MNIST testset

| Methods          | PSNR, dB | Nois realizations |
|------------------|----------|-------------------|
| BM3D             | 24.94    | -                 |
| SDA-N2N-1        | 26.67    | 2 clipped         |
| SDA-N2N-2        | 28.90    | 1 clipped, 1 comp.|
| SDA-N2N-3        | 28.82    | 2 comp.           |
| SDA-SURE-1       | 12.62    | 1 clipped         |
| SDA-SURE w/ comp.| **28.91**| 1 comp.           |
| SDA-SURE         | 28.91    | 1 unclipped       |

### Table 5. PSNR results of denoisers on BSD68 testset

| Methods                        | PSNR, dB |
|--------------------------------|----------|
| BM3D                           | 28.30    |
| BM3D-H                         | 28.40    |
| DnCNN SURE (proposed)          | **28.86**|
| DnCNN SURE (unclipped)         | 28.89    |

5. Conclusion

We proposed a more stable and accurate MC-SURE based training method for deep learning denoisers. Networks trained using our method outperformed conventional state-of-the-art BM3D denoiser and original MC-SURE and yielded comparable results with denoisers trained on clean datasets. Also, we showed that Noise2Noise for i.i.d. Gaussian denoising is a special case of the proposed SURE framework and in case of truncated Gaussian noise, Noise2Noise approach theoretically does not work. Noise2Noise assumes noise to be zero-mean so that it can not be extended to clipped noisy image denoising task. On the other hand, in this work, we further extend our MC-SURE to truncated noise removal task.

Acknowledgements

This work was partly supported partly by Basic Science Research Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Education(NRF-2017R1D1A1B05035810), in part by the Technology Innovation Program or Industrial Strategic Technology Development Program (10077533, Development of robotic manipulation algorithm for grasping/assembling with the machine learning using visual and tactile sensing information) funded by the Ministry of Trade, Industry & Energy (MOFIE, Korea), and in part by a grant of the Korea Health Technology R&D Project through the Korea Health Industry Development Institute (KHIDI), funded by the Ministry of Health & Welfare, Republic of Korea (grant number: HI18C0316).

References

Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., Devin, M., Ghemawat, S., Irving, G., Isard, M., Kudlur, M., Levenberg, J., Monga, R., Moore, S., Murray, D. G., Steiner, B., Tucker, P., Vasudevan, V., Warden, P., Wicke, M., Yu, Y., and Zheng, X. Tensorflow: A system for large-scale machine learning. In Proceedings of the 12th USENIX Conference on Operating Systems Design and Implementation, pp. 265–283, 2016.

Blu, T. and Luijster, F. The SURE-LET Approach to Image Denoising. IEEE Transactions on Image Processing, 16 (11):2778–2786, October 2007.
Theoretical analysis on Noise2Noise using Stein’s Unbiased Risk Estimator for Gaussian denoising

(a) Ground truth  (b) Clipped noisy input  (c) BM3D  (d) SDA-N2N-1  (e) SDA-N2N-2

(f) SDA-N2N-3  (g) SDA-SURE-1  (h) SDA-SURE w/ comp.  (i) SDA-SURE(no clip)

Figure 5. Denoised clipped noisy images from MNIST testset using SDA (noise level $\sigma=25$).

Burger, H. C., Schuler, C. J., and Harmeling, S. Image denoising: Can plain neural networks compete with BM3D? In IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2392–2399, 2012.

Dabov, K., Foi, A., Katkovnik, V., and Egiazarian, K. Image denoising by sparse 3-D transform-domain collaborative filtering. IEEE Transactions on Image Processing, 16(8): 2080–2095, August 2007.

Deledalle, C.-A., Vaiter, S., Fadili, J., and Peyré, G. Stein unbiased gradient estimator of the risk (sugar) for multiple parameter selection. SIAM Journal on Imaging Sciences, 7(4):2448–2487, 2014.

Foi, A. Clipped noisy images: Heteroskedastic modeling and practical denoising. Signal Processing, 89(12):2609–2629, 2009.

He, K., Zhang, X., Ren, S., and Sun, J. Deep residual learning for image recognition. In IEEE conference on computer vision and pattern recognition (CVPR), pp. 770–778, 2016.

Kim, J., Kwon Lee, J., and Mu Lee, K. Accurate image super-resolution using very deep convolutional networks. In IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1646–1654, 2016.

Kingma, D. P. and Ba, J. L. Adam: A method for stochastic optimization. In International Conference on Learning Representations (ICLR), 2015.

Krizhevsky, A., Sutskever, I., and Hinton, G. E. Imagenet classification with deep convolutional neural networks. In Advances in neural information processing systems, pp. 1097–1105, 2012.

Ledig, C., Theis, L., Huszár, F., Caballero, J., Cunningham, A., Acosta, A., Aitken, A. P., Tejani, A., Totz, J., Wang, Z., et al. Photo-realistic single image super-resolution using a generative adversarial network. In IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 105–114, 2017.

Lehtinen, J., Munkberg, J., Hasselgren, J., Laine, S., Karras, T., Aittala, M., and Aila, T. Noise2Noise: Learning image restoration without clean data. In International Conference on Machine Learning, volume 80, pp. 2965–2974, 2018.

Long, J., Shelhamer, E., and Darrell, T. Fully convolutional networks for semantic segmentation. In IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 3431–3440, 2015.

Mao, X., Shen, C., and Yang, Y.-B. Image restoration using very deep convolutional encoder-decoder networks with symmetric skip connections. In Advances in neural information processing systems, pp. 2802–2810, 2016.

Martin, D., Fowlkes, C., Tal, D., and Malik, J. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In IEEE International Conference
Theoretical analysis on Noise2Noise using Stein’s Unbiased Risk Estimator for Gaussian denoising

- Ploetz, T. and Roth, S. Benchmarking denoising algorithms with real photographs. In IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2750–2759, 2017.

- Ramani, S., Blu, T., and Unser, M. Monte-Carlo Sure: A Black-Box Optimization of Regularization Parameters for General Denoising Algorithms. IEEE Transactions on Image Processing, 17(9):1540–1554, August 2008.

- Redmon, J. and Farhadi, A. YOLO9000: Better, Faster, Stronger. In IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 6517–6525, 2017.

- Ren, S., He, K., Girshick, R., and Sun, J. Faster R-CNN: Towards real-time object detection with region proposal networks. In Advances in Neural Information Processing Systems, pp. 91–99, 2015.

- Soltanayev, S. and Chun, S. Y. Training deep learning based denoisers without ground truth data. In Advances in Neural Information Processing Systems, pp. 3261–3271, 2018.

- Stein, C. M. Estimation of the mean of a multivariate normal distribution. The Annals of Statistics, 9(6):1135–1151, November 1981.

- Ulyanov, D., Vedaldi, A., and Lempitsky, V. Deep image prior. In IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 9446–9454, 2018.

- Van De Ville, D. and Kocher, M. Sure-based non-local means. IEEE Signal Process. Lett., 16(11):973–976, 2009.

- Van De Ville, D. and Kocher, M. Nonlocal means with dimensionality reduction and sure-based parameter selection. IEEE Transactions on image processing, 20(9):2683–2690, 2011.

- Vincent, P., Larochele, H., Lajoie, I., Bengio, Y., and Manzagol, P. A. Stacked denoising autoencoders: Learning Useful Representations in a Deep Network with a Local Denoising Criterion. Journal of Machine Learning Research, 11:3371–3408, December 2010.

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

- Zhang, K., Zuo, W., Chen, Y., Meng, D., and Zhang, L. Beyond a Gaussian Denoiser: Residual Learning of Deep CNN for Image Denoising. IEEE Transactions on Image Processing, 26(7):3142–3155, May 2017.