Noise2Noise Verification On Independent But Non-Zero Mean Noise

Yiming Li, ECE undergraduate [1]

Abstract

Noise2Noise is a method to restore corrupted images by learning from only the corrupted samples, yielding similar or even exceeding performance over training with clean data. The application of Noise2Noise method heavily relies on the expectation of the sample noise. Despite the ideal denoising capability for most zero-mean noise, it is argued that Noise2Noise method does not apply to non-zero mean signal-independent noise. A noise with a non-zero mean may alter the overall pixel value of the image, and therefore should jeopardize the performance of Noise2Noise. To verify this argument, in this paper, the Noise2Noise method is re-implemented and tested against standard 0-mean Gaussian noise removal, text removal and non-zero mean Gaussian noise removal. (Su et al., 2020)

1. Introduction

Denoising is a process of reconstructing the original image by removing noise. CNNs with prior knowledge via regular term of loss function is a common method in image denoising. Such supervised learning method aims to find the optimal mapping from corrupted samples to clean ones. The optimization process is achieved by training a regression model,

$$\underset{\theta}{\operatorname{argmin}} \sum_i L(f_\theta(\hat{x_i}), y_i), \tag{1}$$

Where f_θ is a parametric family of mappings under the loss function L.

Training a CNN denoising model usually requires large amounts of pairs of corrupted inputs and clear targets $(\hat{x_i}, y_i)$ to minimize the empirical rick. In practice, obtaining a large amount of high quality training targets is often very challenging due to the limits on radiation exposure and imaging time, in addition to the complications of voluntary (e.g. body) and involuntary motions (e.g. respiratory and muscle relaxation) (Chan et al., 2019). Without the need of clean targets, learning to restore an image using only noisy images in the Noise2Noise model has shown good promise for different noise models such as Gaussian, Poisson, multiplicative Bernoulli, text and random-valued impulse. However, most independent noise is assumed to have a mean of zero. The application of Noise2Noise method is heavily based on the expectation of the sample noise. A hypothesis was made that this method could not apply to non-zero mean signal-independent noise.

In this work, the fundamental principle of Noise2Noise and reasons that it may fail for non-zero mean signal-independent noise are illustrated in section 2. Section 3 first verifies Noise2Noise's application on a common zero-mean noise. Then to verify the hypothesis, Noise2Noise is tested against two types of non-zero mean independent noise. The summary of the results and afterthoughts are documented in section 4.

2. Theoretical Background

Image denoising is to map the corrupted observation to the clean target. For a corrupted input $\hat{x_i}$, and a clean target y_i, The goal for the training process is to look for the optimal set of configuration θ that achieves

$$f_\theta(\hat{x_i}) = y_i \tag{2}$$

where f_θ is a parametric family of mappings (CNNs in this case). For a supervised learning method, the common strategy is to minimize the total deviation of the mapping output from the ground truth value. The deviation can be computed by some loss function L:

$$\sum_i L(f_\theta(\hat{x_i}), y_i) \tag{3}$$

Minimizing the total loss over θ we get (1). For the L_2 loss function, $L(x, y) = (x-y)^2$, the minimization converges to a point where the value of x is the same as the expectation of y. This implies that if the ground truths y are replaced with random numbers that have the same expectation as y, the minimization process and configurations would remain unchanged. This further implies that, for image denoising, the clear ground truth targets for training may be replaced with noise-corrupted samples and the training outcome would not be affected as long as the expectation of the corrupted samples does not vary from the ground truth expectation.

$$\underset{\theta}{\operatorname{argmin}} \sum_i L(f_\theta(\hat{x_i}), \hat{z_i}), \tag{4}$$

where $\hat{z}_i$ is the corrupted sample. The outcome of (4) would be a close estimate to that of (1) if the expectation of z is close to y. Given infinite data, we have $\mathbb{E}[z] = \mathbb{E}[y]$, both outcomes would be empirically the same.

Based on the condition of Noise2Noise method, images corrupted with a zero mean noise should be qualified to serve as the target images. However, samples with a non-zero mean noise would introduce a bias to the expectation. Once the expectation of noisy images deviates from the expectation of clean targets, equation (4) will not hold. Therefore, it is claimed in this work that Noise2Noise method could not apply to independent non-zero mean noises.

3. Practical Experiment

The original Noise2Noise method is proposed and implemented by NVlabs (Lehtinen et al., 2018). The system is written in tensorflow 1 which is not compatible to the most recent environment. A similar implementation of Noise2Noise by yu4u (Uchida, 2019) was used as the baseline model. All the data used in the training and validation are derived from ImageNet (Russakovsky et al., 2015) and the Berkeley Segmentation Dataset (Martin et al., 2001). The local environment is set up with Python 3.8, tensorflow 2.4.1, keras 2.4.3, CUDA 8.0 and cuDNN 11.0. Using the pretrained denoising models, one trained with Gaussian noise samples and Gaussian noise targets (referred as N2N), and the other trained with Gaussian noise samples and clean targets (referred as clean), the denoising outcome of yu4u on Set14 image set is presented in figure 1 and 2. A zero mean Gaussian noise with a standard deviation ranging from 0 to 30 is added to every image from Set14 dataset. The image quality is measured by the Peak signal-to-noise ratio (PSNR) which is the ratio between the maximum possible power of a signal and the power of corrupting noise.

$$PSNR = 20 * log_{10}(\frac{L-1}{\sqrt{MSE}}) \tag{5}$$

where L is the number of maximum possible intensity levels in an image (i.e 256). MSE is defined by

$$MSE = \frac{1}{mn}\sum_{i=0}^{m-1}\sum_{j=0}^{n-1}(O(i,j) - D(i,j))^2 \tag{6}$$

Where O and D represent the matrix data of original image and degraded image; m, n, i, j represent the numbers of rows and columns of pixels and index of that row and column of the image respectively.

For each noise level, the psnr values yielded from denoising all testing images are averaged. Figure 1 shows the PSNR vs. noise level curve of the N2N model, and figure 2 shows the curve of the clean model. The results indicate a performance similarities between traditional denoising and Noise2Noise denoising for a 0 mean Gaussian noise.

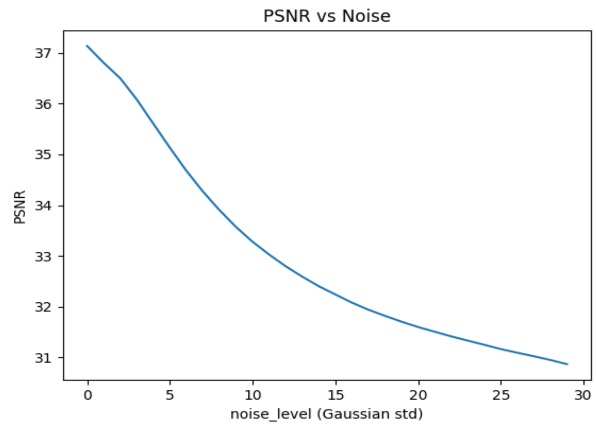

Figure 1. PSNR vs. Noise of Set14 using model trained with Gaussian noise and noise target

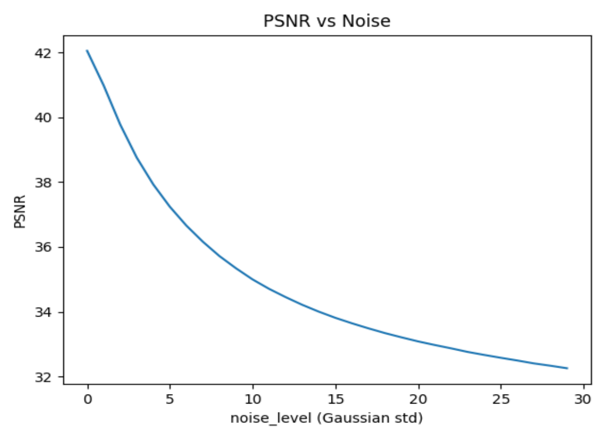

Figure 2. PSNR vs. Noise of Set14 using model trained with Gaussian noise and clean target

3.1. Zero Mean Gaussian

Gaussian noise is a common image noise that can be found in various sources such as CT images (Wu et al., 2019). To further verify the Noise2Noise method, a Gaussian denoising model is retrained with Set 291 as the training set and Set 14 as the validation set. The model adopts super-resolution residual neural network (SRResNet) as the architecture. The training samples are used to generate a source set and a target set. The source set contains all the images from the training set but corrupted with a Gaussian noise. This Gaussian noise has a zero mean and a uniformly distributed standard deviation ranging from 0 to 50. The standard deviation varies because in reality the noise level is not constant (Hasan et al., 2021). For the traditional CNN clean model, the target set contains clean images from the training set. For Noise2Noise, the target set contains images

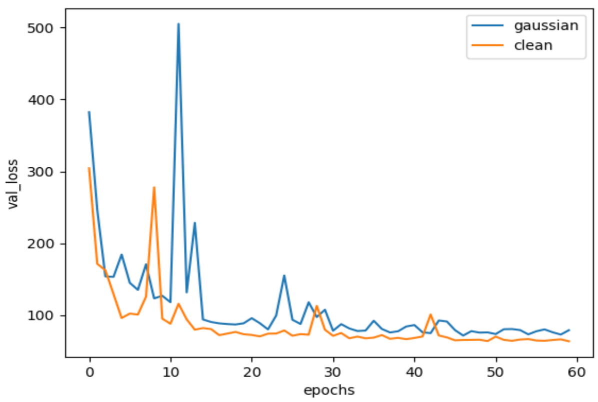

Figure 3. Loss over epochs for N2N (gaussian) and clean (clean) model

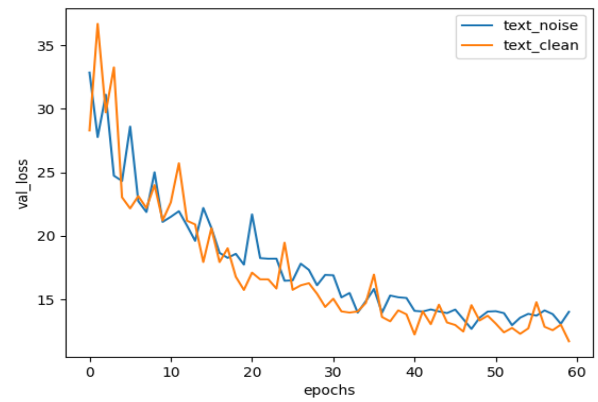

Figure 5. Loss over epochs for N2N (text_noise) and clean (text_clean) model

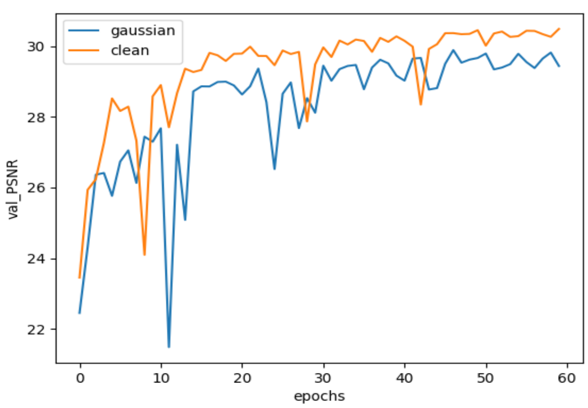

Figure 4. PSNR over epochs for N2N (gaussian) and clean (clean) model

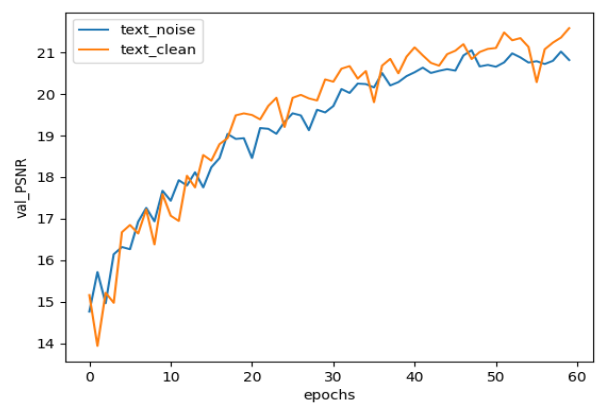

Figure 6. PSNR over epochs for N2N (text_noise) and clean (text_clean) model

corrupted with the same Gaussian noise as the source set. To validate the denoising outcomes, Set 14 image set is corrupted by a Gaussian noise and fed into both models. This Gaussian noise has a zero mean and a standard deviation of 25. To visualize the training history, the MSE loss and PSNR values across all training epochs are plotted in figure 3 and 4.

It can be observed that the model trained with clean targets performs slightly better than the model trained with noise targets. The clean model overall has a lower loss and a higher PSNR value than the Gaussian model does across most epochs.

An example of the denoising outcomes from both the Noise2Noise Gaussian model and traditional CNN clean model is shown in figure 5. The PSNR values are close for both denoising outcomes. Noise2Noise outcome has a PSNR of 30.53 (dB) and CNN outcome has a PSNR of 30.62 (dB) compared to the original picture. The visual quality of both outcomes are satisfying. Noise2Noise method can perfectly apply to zero mean Gaussian noise.

3.2. Text Removal

Additive Text is a common source of image noise. Previously, a 0 mean Gaussian noise is added to every pixel of the image, thus the mean of the corrupted image remains the same. Additive text however does not affect every single pixel. Instead, it is a signal-independent noise that could randomly appear at any places of the image. Such noise consists of a large and varying number of random strings, and these strings can overlap on each other.

To simulate this noise, strings with a random length of 5 to 10 of random ascii letters and digits are generated. The

Figure 7. Denoising outcomes examples for 0 mean Gaussian noise: Left to Right: Original, Corrupted, N2N, CNN

Figure 8. Denoising outcomes examples for additive text: Left to Right: Original, Corrupted, N2N-L_1, CNN, N2N-L_2

font is set to *Hershey Simplex*. The font scale, thickness, color, and string location are also randomized. These texts are directly added to the images to create corrupted samples. The expectation of text corruption depends on its color, which is a set of 3 uniformly distributed random integer between 0 and 255. The average is $(128, 128, 128)$, a gray color. Thus, the mean of the image is very likely to change with the addition of text.

Figure 6 showcases the denoising outcomes of a zebra. For the L_2 loss function, $L(x, y) = (x - y)^2$, the minimization converges to a point where the value of x is the same as the expectation of y. The fifth zebra is the denoising outcome of N2N method using L_2 norm. Compare it with the first zebra, the original image, there is a noticeable difference in the visual. The N2N-L_2 outcome appears to be blurrier due to the certain amount of greyness caused by the additive text added to the picture. More specifically, the outcome would incorrectly tend towards a linear combination of the right answer and the average text color (Lehtinen et al., 2018). This is similar to the issue of outliers in linear regression. The square operation of L_2 loss amplifies the individual error terms. When the relative significance is weighted between outliers and the normal data points, the outliers will have a large contribution to the error. To minimize the total loss, the optimization solution will be driven by the outliers to compensate for the larger error (Chan, 2021). One way to circumvent the issue of outliers, or in this case the pixel value bias introduced by text, is to replace the squared error by the absolute error. Therefore, L_1 loss is a more preferable loss function to L_2 loss. For L_1 loss, $L(x, y) = |x - y|$, the minimization converges to the optimum when x approaches the median of y. Unlike the Gaussian noise which deteriorates every pixel, additive text only affects a small portion of the pixels so that the majority of the pixels remain their color. The median of the pixel values will very likely remain the same as well, and therefore L_1 loss is suitable for text removal.

The third and the fourth zebra are the denoising outcomes of N2N and CNN clean model using L_1 loss. Visually speaking, both outcomes to a greater extent preserve most of the pixel colors comparing to the firth one. From the third to the fifth zebra, the PSNR values are 32.27(dB), 32.54(dB), and 28.31(dB). Figure 5 and 6 visualize the training history for N2N and CNN clean model both using L_1 loss. As shown, the performance of noise2noise model is very close to the performance of the clean model. The noise2noise method still remain a good applicability for text removal.

3.3. Non-Zero Mean Gaussian

In the previous verification of text removal, a reasonable amount of additive text would not change the majority of

Figure 9. Denoising outcomes examples for 100-mean Gaussian noise: Left to Right: Original, Corrupted, N2N, CNN

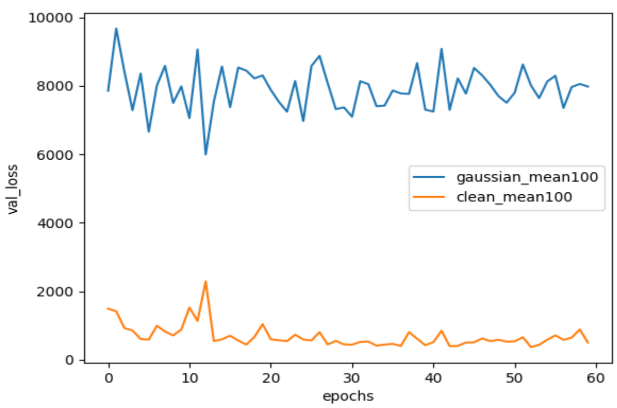

Figure 10. Loss over epochs for N2N (gaussian_mean100) and clean (clean_mean100) model

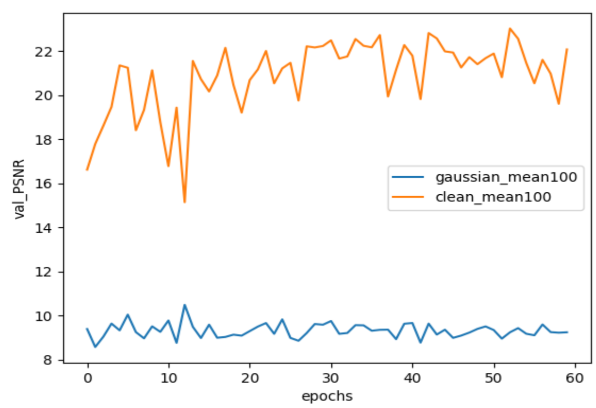

Figure 11. PSNR over epochs for N2N (gaussian_mean100) and clean (clean_mean100) model

the pixel values, therefore it only introduces a small amount of bias to the overall data statistics. However, a non-zero mean Gaussian noise affects on every individual pixel and changes the average value of an image. To demonstrate Noise2Noise applicability to this noise, a denoising model is trained with Set 291 as the training set and Set 14 as the validation set. Similar to the zero mean Gaussian denoiser, the additive Gaussian noise has a standard deviation ranging from 0 to 50. Instead of a mean of 0, this Gaussian noise has a mean of 100. For Noise2Noise model training, the target set contains images crrupted with the same Gaussian noise as the source set. For CNN clean model training, the target set contains the unbiased clean images from the training set. The loss and PSNR value across the epochs in the training history are presented in Figure 8 and 9. Unlike the previous 0-mean Gaussian plots, for 100-mean Gaussian noise, there exists an apparent performance gap between Noise2Noise and the clean CNN denoisers. Noise2Noise has much higher loss values and lower PSNR values across all epochs.

Figure 7 showcases examples of the denoising outcomes of the N2N denoiser and the CNN clean denoiser. The second coastguard is corrupted by a 100-mean Gaussian noise with a standard deviation of 25, and it is fed into the N2N and CNN denoiser as an input. The third and the fourth coastguards are the denoising outcomes. Visually speaking, the N2N is able to smoothen the image and alleviate the blurriness caused by the electronic circuit noise (the primary cause of 0-mean Gaussian noise) (Mafi et al., 2019). However, the N2N-denoised outcome is considerably brighter than the original image because of the noise's 100-mean. Statistically speaking, the N2N-denoised outcome has a PSNR value of 27.42(dB) and the CNN-denoised outcome has a considerably higher PSNR value of 30.99(dB). As regard to the image mean, the original image has a mean of 113.34; the N2N-denoised outcome has a mean of 195.00 whereas the CNN-denoised outcome has a mean of 117.75 which is very close to the original image mean. Because the N2N denoiser is trained with respect to the 100-mean Gaussian corrupted targets, both sides of the N2N denoiser mapping have the same image mean, and thus the mapping function will not attempt to reduce the brightness. In contrast, the CNN clean denoiser trained with clean targets is able to correct the false mean value of the input images. At the training process, with clean images as targets, the CNN clean denoiser mapping will attempt to map corrupted images with a false mean to clean images with the original mean. Thus it's able to reduce the mean of the validation image by roughly 100, reconstructing the image in a more true-to-original way.

4. Discussion

It is shown that recovering signals under standard zero-mean Gaussian corruption without observing clean signals is achievable with Noise2Noise method. Without the direct knowledge of clean image statistics, Noise2Noise method relies on the stability of image pixel expectation under the corruption of noise and aims to retrieve the true statistics of clean images from the noisy samples. However, non-zero mean introduces a bias to the image pixels and breaks this stability of image pixel expectation, and thus it may jeopardize the performance of Noise2Noise method. For a certain type of non-zero mean signal-independent noise, additive text, even though the image pixel expectation is changed, it is still possible to recover signals with only noisy observations by tuning to the median of the observations which remains unchanged under text corruption. Nevertheless, signals corrupted by a non-zero mean Gaussian noise does not remain the clean image statistics. Noise2Noise method therefore is not able to correct the false mean of the noisy inputs and thus fails the validation.

References

C. Chan, J. Zhou, L. Yang, W. Qi, and E. Asma. Noise to noise ensemble learning for pet image denoising. In *2019 IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC)*, pages 1–3, 2019. doi: 10.1109/NSS/MIC42101.2019.9059779.

S. H. Chan. *Introduction To Probability For Data Science*. 2021.

A. M. Hasan, M. R. Mohebbian, K. A. Wahid, and P. Babyn. Hybrid-collaborative noise2noise denoiser for low-dose ct images. *IEEE Transactions on Radiation and Plasma Medical Sciences*, 5(2):235–244, 2021. doi: 10.1109/TRPMS.2020.3002178.

J. Lehtinen, J. Munkberg, J. Hasselgren, S. Laine, T. Karras, M. Aittala, and T. Aila. Noise2noise: Learning image restoration without clean data. *CoRR*, abs/1803.04189, 2018. URL http://arxiv.org/abs/1803.04189.

M. Mafi, H. Martin, M. Cabrerizo, J. Andrian, A. Barreto, and M. Adjouadi. A comprehensive survey on impulse and gaussian denoising filters for digital images. *Signal Processing*, 157:236–260, 2019. ISSN 0165-1684. doi: https://doi.org/10.1016/j.sigpro.2018.12.006. URL https://www.sciencedirect.com/science/article/pii/S0165168418303979.

D. Martin, C. Fowlkes, D. Tal, and J. Malik. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In *Proc. 8th Int'l Conf. Computer Vision*, volume 2, pages 416–423, July 2001.

O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei. ImageNet Large Scale Visual Recognition Challenge. *International Journal of Computer Vision (IJCV)*, 115(3):211–252, 2015. doi: 10.1007/s11263-015-0816-y.

J.-W. Su, H.-K. Chu, and J.-B. Huang. Instance-aware image colorization, 2020.

Y. Uchida. noise2noise. https://github.com/yu4u/noise2noise, 2019.

D. Wu, K. Gong, K. Kim, X. Li, and Q. Li. Consensus neural network for medical imaging denoising with only noisy training samples. In D. Shen, T. Liu, T. M. Peters, L. H. Staib, C. Essert, S. Zhou, P.-T. Yap, and A. Khan, editors, *Medical Image Computing and Computer Assisted Intervention – MICCAI 2019*, pages 741–749, Cham, 2019. Springer International Publishing. ISBN 978-3-030-32251-9.