Final report (Option 1)

Course name: Multimedia Processing Technique (DD026_1594)

Project number: Option 1-6

Project name: Survey the paper: **Noise2Noise: Learning Image Restoration without Clean Data**

Paper link: https://arxiv.org/abs/1803.04189

Student name: Son hayeon. 201811300

Github of project: https://github.com/abc

Report:

1) Introduction

Signal reconstruction in damaged or incomplete measurements is an important subfield of statistical data analysis.

Recent traditional methods of deep neural networks are mapping from clean images, which is often difficult or quite inefficient to obtain this "clean image."

Thus, in this paper, we present a learning method for restoring images without clean data.

2) Theoretical background

Unreliable room temperature measurements (y1, y2...)Let's assume that there is

$$\underset{z}{\operatorname{argmin}} \, \mathbb{E}_y\{L(z, y)\}. \quad (2)$$

A common strategy for measuring the true unknown temperature is to find z as above (z = the smallest mean deviation of the measurements according to the loss function L)

$$z = \mathbb{E}_y\{y\}. \quad (3)$$

For L2 loss (L(z, y) = (z-y)^2), this minimum value z can be found in the arithmetic mean of the observations as above.

The L1 loss (L(z, y) = |z-y|), the sum of the absolute deviations, can be found to be the optimal value from the median value of the observation.

From a statistical point of view, summary estimation using this common loss function can be viewed as ML estimation by interpreting the loss function as a negative log.

$$\underset{\theta}{\operatorname{argmin}} \mathbb{E}_{(x,y)}\{L(f_\theta(x),\, y)\}. \tag{4}$$

Neural network regression analyzer training is a generalization of procedures for estimating these z-values. As above, we observe the form of general training tasks for a set of input-target pairs (xi,yi) where the network function f(x) is parameterized by theta:.

In practice, if we remove the dependence on input data and use a trivial f_theta that outputs only learned scalars, we work on reducing it to (2). Conversely, in all task training, the packed problem decomposes the sample. We show that the simple operation is equivalent to (4).

Theoretically, the network can minimize this loss by solving the z estimation problem separately for each input sample. Thus, the attribute of the underlying loss is inherited by neural network training.

A typical process of regression training by Equation 1 for a finite number of input-target pairs (xi, yi) is a 1:n mapping, not a 1:1 mapping between the input and target (false) implied by that process. That is, p(y|x) is a very complex distribution of any image consistent with the low resolution x. Using L2 losses to train neural network regressors using training pairs of low-resolution and high-resolution images, the network learns how to output the average of all values that result in spatial blurring for prediction of the network. A considerable amount of work has been done to overcome this trend.

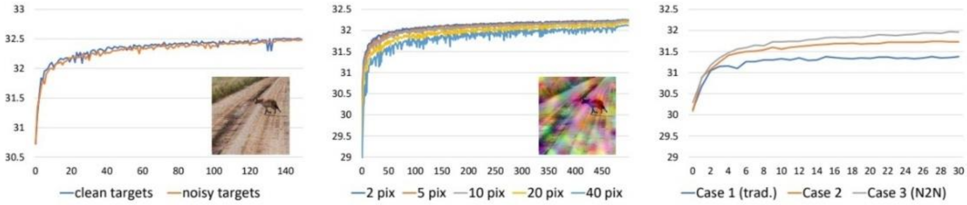
The author of this paper found that this tendency has an unexpected advantage for certain problems. The trivial and unimportant property of L2 minimization is that if the goal is replaced by an arbitrary number consistent with the author's goal according to expectations, fewer estimates remain unchanged.

In many image restore operations, the ideal result value of corrupted input data is the clean image we are trying to restore. For example, L1 loss restores the median of the target, which means that neural networks can be trained to recover images with significant (up to 50%) outlier content, which requires only two corrupted images.

The contents of the study present various examples showing that these theoretical functions are actually efficiently feasible.

3) Contents of the study

In 3.1, we study recovering damaged images using Gaussian noise. Since the mean of noise is 0, we recover the mean with learning using L2 loss.



(a) White Gaussian, $\sigma = 25$ **(b) Brown Gaussian, $\sigma = 25$** **(c) Capture budget study (see text)**

(a) (a) In , clean and noisy targets exhibit very similar convergence speed and eventual quality.

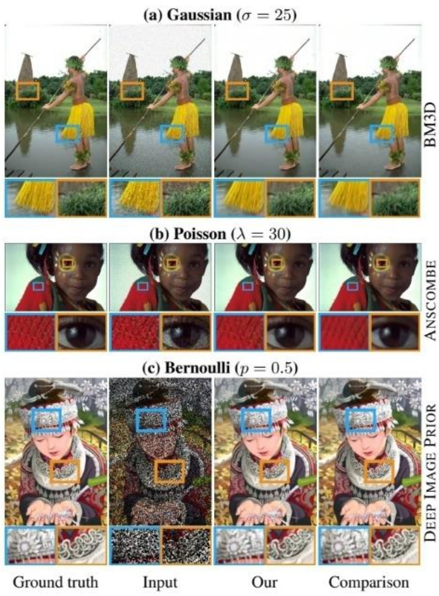
(b) (b) In , we can see that the inter-pixel nose correction is slow but the performance is good.

(c) (c) In , you can see the difference between the budget for the noisy sample and the clean sample.

Previous studies have relied on an infinite number of noisy samples produced by adding artificial composites to turn them into clean images. However, as a result of the study, it was found that the budget could be secured only with finite data and fixed training data.

Through this, it can be seen that in the case of additional Gaussian noise, the damaged target provides not only the same performance but also better performance compared to the clean target. It also recognizes more damage to the same latent clean image in providing two advantages, and it is possible to see more latent clean images even if there are only two corrupted targets.

In 3.2, we experiment with different types of synthetic noise.



Poison noise is the main source of noise in photography, which is more difficult to eliminate because the mean is zero but depends on the signal.

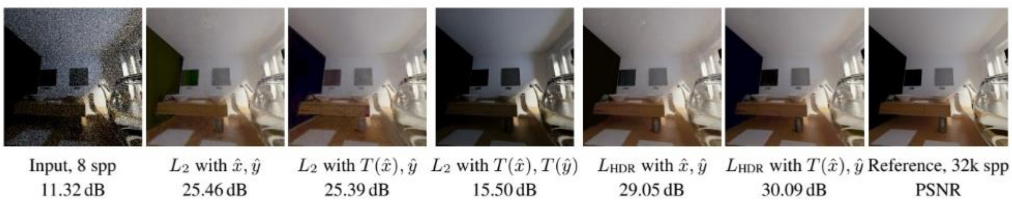
Bernoulin noise is a random mask configuration with a valid pixel value of 1, and a missing pixel value of 0, excluding the loss to prevent the gradient from being reversed in the missing pixel.

In the above figure, we can see the results of the examples for Gaussian, Poison, and Bernoulli noise. As a result of using noisy targets, we can observe similar or better results than using clean targets.

These loss functions also show that it is possible to remove text or random impulse noise from an image.

Figure 3. Removing random text overlays corresponds to seeking the median pixel color, accomplished using the L_1 loss. The mean (L_2 loss) is not the correct answer: note shift towards mean text color. Only corrupted images shown during training.

3.3 shows that Monte Carlo noise can be eliminated as shown in the following picture.

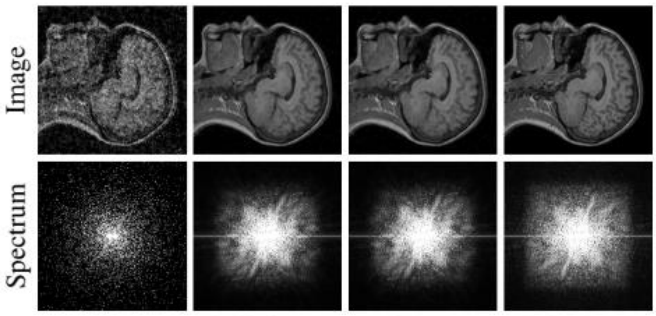


(a) Input (64 spp), 23.93 dB (b) Noisy targets, 32.42 dB (c) Clean targets, 32.95 dB (d) Reference (131k spp)

Monte Carlo rendering is frequently used in virtual environments, resulting in random order of light scattering in scenes connecting light sources and virtual sensors. Noise generated from these events is configured such that the average value is zero by Monte Carlo integrators, but Monte Carlo noise is much more difficult than removing Gaussian noise because there is no clear answer to the distribution.

3.4 studies MRI.

MRI produces stereoscopic images of biological tissue through the Fourier transform (the "k-space") sampling of signals. This k-space sampling is converted into a random process with probability density P(k) and the study is conducted through frequency k.



(a) Input 18.93 dB (b) Noisy trg. 29.77 dB (c) Clean trg. 29.81 dB (d) Reference

Through these studies, it can be seen that results similar to clean targets as in the picture above can be obtained.

4) The conclusion

The author found a way to recover from a simple statistical argument through an image that is not clean. This proved that it could perform as much as or more than using a clean image, and that clean data is not necessarily necessary for noise cancellation.

This suggests that the performance of deep neural networks can be highly resilient and can serve as significant benefits in many areas.