SciencePG
Science Publishing Group

# Similarity Noise Training for Image Denoising

## Abderraouf Khodja, Zhonglong Zheng[*], Yiran He

College of Mathematics and Computer Science, Zhejiang Normal University, Jinhua, China

**Email address:**
201710800010@zjnu.edu.cn (A. Khodja), zhonglong@zjnu.edu.cn (Zhonglong Zheng), he1ran@qq.com (Yiran He)
[*]Corresponding author

**Abstract:** Deep learning has attracted a lot of attention lately, thanks. Thanks to its high modeling performance, technological advancement, and big data for training, deep learning has achieved a remarkable improvement in both high and low-level vision tasks. One crucial aspect of the success of a deep learning-based model is an adequate large data set for fueling the training stage. But in many cases, well-labeled large data is hard to acquire. Recent works have shown that it is possible to optimize denoising models by minimizing the difference between different noise instances of the same image. Yet, it is not a common practice to collect data with different noise instances of the same sample. Addressing this issue, we propose a training method that enables training deep convolutional neural network models for Gaussian denoising to be trained in cases of no ground truth data. More specifically, we propose to train a deep learning-based denoising model using only a single noise instance. With that in mind we develop a non-local self-similarity noise training method that uses only one noise instance.

**Keywords:** Image Denoising, Convolutional Neural Networks, Block Matching, Unsupervised Learning,
Non-Local Self-Similarity

## 1. Introduction

Recently, deep learning based models have pushed computer vision tasks to a new level of performance, making deep learning the go to framework when dealing with images, many breaks throw have been witnessed in high-level tasks like Image Classification [1-3], Object Detection [4-6] and low-level vision like in Image Denoising [7-10], Inpainting [11] Single Image Super Resolution [12] and more. A great emphasis has been placed in developing deeper and more complex architectures to improve modeling performance, as well as improving training technics.

One major reason that enables deep learning models to achieve such success is the availability of large labeled data sets for fueling the training process, and by the same mean achieving high performance, yet large data sets with clean ground truth labels are not always given, and sometimes even not feasible.

Image denoising is a low-level vision problem where deep learning has become extremely popular, it is also a pre-processing step in different computer vision tasks, it aims

to solve an inverse problem of the form $Y = X + v$ where we seek to estimate the clean image $X$ from its noisy measurement $Y$ by reducing the noise perturbation $v$. Across the denoising literature and specially in natural image denoising, it is usual to model the noise perturbation as Additive White Gaussian Noise (AWGN), though real world noise does not match exactly this distribution it is still useful in practice to model it as so.

Two main approaches can be distinguished in the denoising literature, following its chronological development we encounter prior based methods first [13-17], sparsity and non-local self-similarity played a central role in the development of these prior based methods. These priors are usually mathematical models engineered by researchers so as to model the inherent images structure. These methods were the reference choice when dealing with image denoising problems, indeed the quest for best denoising performance was till lately an improvement over the mathematical modeling of the image inner structures. Though providing a well-founded theoretical foundation and stable convergence criteria, these hand-crafted models lakes of modeling performance, as no model can perfectly describe images. This has led to

the adoption of learning-based models. This second stream of methods has become quite popular lately, in this framework the burden of engineering the model has been lifted by learning strategies [7, 8, 10, 18] and has led to state-of-the-art performance in most computer vision tasks.

In a deep learning based framework, two main aspects are critical for the model success, one is the model architecture and second is training data. For training a deep learning model large labeled data set need to be at hand. In terms of image denoising, pairs of (noisy, ground truth) samples are needed to perform the training process by minimizing a loss function (Mean Squared Error (MSE) for example) between the network estimation on the noisy sample and its ground truth (noise free) images. That being said, in many cases ground truth data can be hard to acquire or technically consuming, one example is in medical imaging, where 3D Magnetic resonance imaging (MRI) needs hours of acquisition for a single high quality volume and reducing time acquisition leads to noise perturbation, which is harmful for medical diagnose.

Recently, an interesting property has been observed in the context of image restoration tasks [19], in this work it is shown that the training process of various image restoration tasks including image denoising can be robust to normal perturbations in ground truth data, meaning that if the ground truth data used during training had to be altered by a zero mean Gaussian noise then this would not significantly harm the denoising performance, it has been also shown that in many cases we can achieve similar denoising performance. Said differently we can train a deep learning denoising model by minimizing the loss of two different noise degradation of the same sample and still achieve denoising outcome. This sort of behavior is rather appealing, and motivate us to further explore other training scenarios with no ground truth image.

In this paper we propose to approach the issue of training deep learning based denoising models in case of no ground truth data set, in contrast with noise2noise training our method relays on using only single noise degradation samples. We propose to do that by joining noise2noise training with the non-local self-similarity image prior. To that end, we design a group patch based convolutional neural network (CNN) model that suits the need of non-local self-similarity and noise2noise training method [19], we discuss the training process which raises a certain number of difficulties and propose adequate approaches to deal with them.

## 2. Related Works

Lately different attempts have been conducted in terms of no ground truth dataset training, in [19], it has been pointed out that deep neural networks can be trained without the need of ground truth data by mapping noisy to noisy samples, in [18] a generative adversarial network has been proposed for forming pairs of training data's, and has led to impressive

results in image blind denoising. Others like [20] have chosen to introduce a statistical noise prior by modifying the training objective function using Stein's unbiased risk estimate (SURE) estimator.

In [19], it is assumed to have pairs of noisy data, which is not really common, usually it is not relevant to acquire couples of noisy data, in contrast we propose to use single instance of noise data, and form our pairs from similar patches. In in [18] assumes statistical prior knowledge of the noise induced by a SURE noise estimation, here we assume no statistical noise prior knowledge.

## 3. Our Method

Our goal is to design a no ground-truth training method for image denoising, most existing deep learning methods namely supervised learning relay exclusively on training the model by minimizing the model prediction with its corresponding label or ground truth,

$$\ell(\Theta) = \frac{1}{N} \sum_{n=1}^{N} \|Net(y_i, \Theta) - x_i\|_F^2 \qquad (1)$$

It is obviously the default approach to adopt if available. Here, $Net$ is the model to be optimized, $\Theta$ the model's parameters, $N$ the number of training samples and the (noisy, clean pairs) $\{(y_i, x_i)\}_{i=1}^{N}$. From [19] we know that in the context of image restoration training is not only restricted to ground truth data, and noisy target can also achieve satisfying denoising performance, this property is highly interesting since it remarkably changes our perceptions of how learning models can achieve specific tasks. But in case where no multiple degradations of the same sample are available it is still not clear how can we use this property to our advantage.

Indeed, it is rare to find datasets that include pairs of noisy data, in fact, the most common case is to find data of only single sample noise instance captions, here the noise2noise training method proposed in [19] while being highly relevant, does not meet an optimal practical training requirement, since the noise2noise training method relays on the availability of multiple degradations of the same sample which makes impractical if else. It would be advantageous to find a way train exclusively on noisy data, yet, a learning model needs an objective function with training pairs for optimize the model parameters and in this case non-local self-similarity suits perfectly our need.

Natural images exhibit redundancies across different spatial position in the image, it is thus possible to extract training pairs from the given data using similar patches, which enables us to train a denoising model using only corrupted data with no corresponding ground truth versions based on a noise2noise training method.

Motivated by the above reasoning, we combine the non-local self-similarity prior with noise2noise training method, we also design an appropriate CNN architecture for that matter, our training model is composed of two networks.
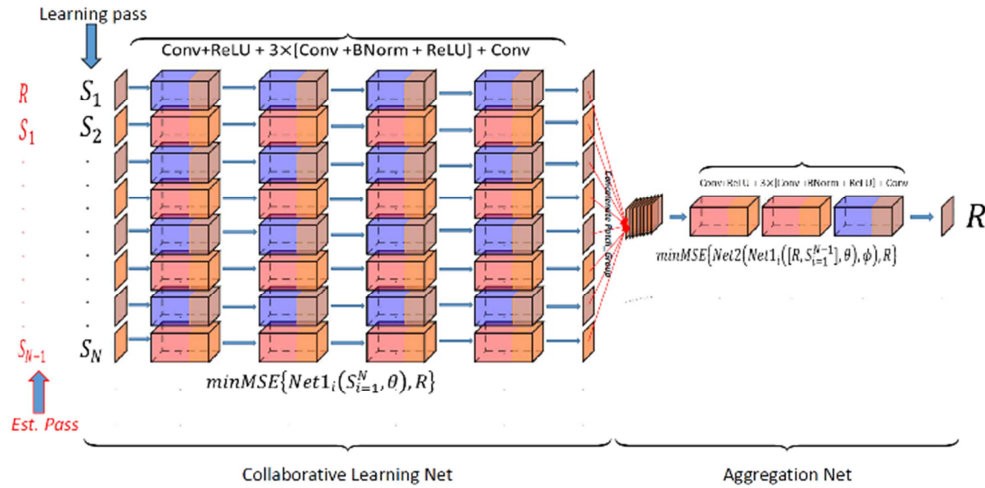
**Figure 1.** *Overall pipeline of the two networks.*

This being said we can proceed to the training process. First, let us dive into the detail of our model for similarity noise training method. In the following subsections, we will present the details of our training and implementation, we take inspiration from the way Block-matching and 3D filtering (BM3D) [14] exploit non-local self-similarity and we propose a patch group based CNN model; a first network takes as inputs groups of N similar patches are formed after performing a block matching on the noisy images (with respect to a reference patch in research window of size W); these N patches are then fed to a first CNN and denoised jointly, the resulting estimates are afterward fed to a second CNN, this second network would provide the final estimation of the clean reference patch, Figure 1 shows a diagram of the two networks architecture.

If we would make the analogy with BM3D, the first network would be the equivalent of the collaborative filtering. Though here, we employ it in terms of learning, our assumption is that equivalently to collaborative filtering that aims the enhance denoising performance by filtering together groups of similar patches, we can expect our training process and by the same mean the denoising performance to gain from training and denoising groups of similar patches together. We name this network Collaborative Learning-Net (CL-Net), the second network aims to aggregate the group of similar patches and we call it Aggregation-Net (Ag-Net), more details about the CL-Net and Ag-Net in subsection 3.2.1 and 3.2.2.

### 3.1. Block Matching

In the settings of our training, we choose the patch size to be $(20, 20)$, most state-of-the-art methods that implements a non-local self-similarity prior uses a small patches size for block matching, generally $(7, 7)$, in the context of CNN and for the learning purposes, a larger patch size is usually needed, we have found empirically that $(20, 20)$ achieves a good balance between learning size and block matching error, typically the Euclidean distance (MSE) is used as similarity measure, similar patches are collected in a $(27, 27)$ window

centered around the reference patch. Our patch groups will be formed of $N = 9$ patches including the reference patch we will refer to a patch as $\{S_i\}, i = 1, N$ and $S_1 = R$ the reference patch, we have found imperially that increasing the of patch in the groups is produces denoising artifacts.

### 3.2. Model Architecture

Designing a suitable network architecture for the task at hand is a crucial step for the model success, here we do seek an architecture that both suits a patch group formulation with a noise to noise driven training, state-of-the-art CNN [7] denoising models could achieve state-of-the-art denoising performance with a $(3, 3)$ filter size, the same has been previously indicated by Simonyan et al. [21]; also the network depth plays an essential role in determining the receptive fields size, in [7] it has been shown that for a $(3, 3)$ filter size CNN the depth of the networks correlates with the receptive fields size, and convolutional layers CNN would result in a $(2d + 1, 2d + 1)$ receptive field.

Since working in patch based model we choose the depth of the network to be rather small, a 5 convolutional layer deep network, Rectifier Liner Unit ReLU $max(z, 0)$ are adopted as activations functions. Batch Normalization shows to enhance performance and here again we adopt Batch Normalization in our network.

It is important to distinguish between the Collaborative Learning-Net pipeline and the Aggregation-Net one, we choose to perform convolutions for the Collaborative Learning-Net in a filter group manner, 8 groups of 64 filters per layer for each channel, this means that though passing through the same network the channels are actually processed separately, but still working jointly, this is because we choose to map each one of their estimations to the reference patch. The reason for adopting this strategy is that when adopting a combined forward pass, the Collaborative Learning-Net ends up with all the patches in the group producing the same outcome (estimations of the reference patch) and loses their distinctive structures, and that do not satisfy our needs for aggregation, it is thus preferable that the Collaborative

Learning-Net learns to estimate denoised patches while preserving the patches structures.

To summarize:

*Collaborative Learning-Net:*

1. Feature Extraction: Convolutions of 8 filters groups of size (3,3,1,64) + ReLU activation function
2. Feature processing: Three blocks of Convolution of 8 filters groups of size (3,3,64,64) + Batch Normalization + ReLU
3. Deconvolution: Convolution of 8 filters groups of size (3,3,64,1)

*Aggregation-Net:*

1. Feature Extraction: Convolutions of filters of size (3,3,9,64) + ReLU
2. Feature processing: Convolution filters of size (3,3,64,64) + ReLU
3. Deconvolution: Convolution of filters of size (3,3,64,1)

### 3.2.1. Collaborative Learning

The Collaborative Learning-Net is designed to perform two different tasks simultaneously. In one hand, our network learns its denoising function; as we are training this network to map groups of similar patches inputs to their reference noisy one. This follows our previous assumption on similarity noise to noise training method, Figure 2 representing MSE loss across different epochs using only one noisy image verifies our assumption, we can clearly see a convergence toward a denoised outcome. We can also see how training our networks by minimizing the MSE between the prediction of our model on groups of noisy similar patches to their reference noisy patch converges towards minimizing the MSE of the same groups with their reference ground truth patch, this proves that our noise similarity training method is achieving the expected denoising behavior.

Moreover, we can see a high correlation between ground truth target loss and the noisy target loss which again shows the reliability of the method. On the second hand while learning to restore the patch groups it also learns to work collaboratively, this is mainly due to the fact that during the learning process all the N similar patches referred as $(S_i, i = 2, N)$ in each group are trained to map the same reference patch

$$minMSE\{[CL\_Net(S_{i=2}^N, \Phi)), R\} \qquad (2)$$

and thus, collaborates to its final aggregation, note that this formulation includes only noisy samples from the noisy image.

### 3.2.2. Aggregation

The Aggregation-Net works at producing the final estimate of the clean reference patch; previously the Collaborative Learning-Net did a good job at denoising, and outputs patch groups that approaches the references patches; but still these patch group's estimations are only linked by their objective target (i.e., mapping the reference noisy patch) and did not yet produce a clean estimation of the reference patch, the Aggregation-Net that takes the N similar patches and pass them jointly into one channel in the Aggregation-Net, here again we employ a noise mapping formulation, meaning that the Aggregation-Net will be trained to map to the reference noisy patch.

$$minMSE\{Ag\_Net\left(\left[CL_{Net(S_{i=1}^{N-1}, \Phi)}, G_{mean}\right], \Psi\right), R\} \qquad (3)$$

CL-Net and Ag-Net denote the Collaborative Learning-Net and the Aggregation-Net and $\Theta, \Psi$ their respective parameters. Note that when the Collaborative Learning-Net is being optimized, it is fed with similar patches without their reference patch, during our experiments we have noticed that providing the CL-Net with the reference patch in the patch group input tends overfit our model since then the network would just learn to copy noisy patches. Once the CL-Net network converges, it is used to produce the inputs of the second network, this time including the reference patch. The group mean of the estimated patches which is defined as

$$G_{mean} = \frac{1}{N}\sum_{i=1}^{N-1} S_i \qquad (4)$$

$G_{mean}$ is also passed to the Aggregation-Net we found that this help the Aggregation-Net to better estimate the clean reference patch by integrating a basic estimation of it of the estimated patches.
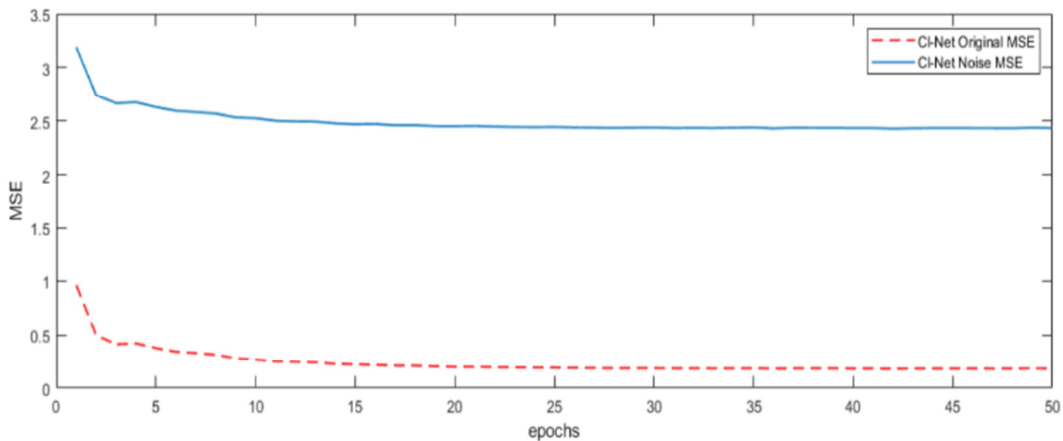


***Figure 2.*** *Mean squared error during training on pairs of similar patches from a single image.*

## 4. Experiments

Here we present the experiment conducted following our similarity noise training method, evaluation and performance. To analyze the training process and denoising result we conduct the training on a test set of 12 widely used images, Figure 3 shows the test set; also the BSD500 [22] has been used to train on a larger scale and study the effect of big data set for the training, 400 training images are cropped to ( 180, 180) and AWGN is added before performing block matching, and groups of similar patches are formed, flips and rotation for data augmentation are performed.



*Figure 3. Test images used for training and testing.*

We consider three training data set, one image training data set, test dataset training and external data set training (BDS500), Table 1 show the denoising performance on for the three datasets. It is interesting to see that training on a single image achieve the best result in terms of PSNR, followed by test set (12 Set) training and last is the external set (BSD500) training. Obviously, the training can better adapt to the given data and thus achieve better denoising than the model trained on a different dataset. This encourages us to better develop the model for practical use in cases of no ground truth data.

*Table 1. The average PSNR (DB) results of different training datasets on the 12 images set. The best results are highlighted in Red first and Blue second with noise level 25.*

| TestSet | Single image training | Set12 training | BSD500 training |
|---|---|---|---|
| Cameraman | 26.75 | 26.33 | 26.45 |
| House | 31.95 | 31.3 | 30.98 |
| Peppers | 28.51 | 28.4 | 28.28 |
| Starfish | 27.09 | 27.06 | 27.28 |
| Monarch | 27.97 | 27.39 | 27.43 |
| Jetplane | 26.58 | 26.37 | 26.65 |
| Parrot | 26.95 | 26.55 | 26.69 |
| Lena | 30.62 | 31.63 | 30.78 |

| TestSet | Single image training | Set12 training | BSD500 training |
|---|---|---|---|
| Barbara | 29.83 | 28.56 | 26.5 |
| Boat | 28.83 | 28.94 | 28.6 |
| Man | 28.87 | 29.03 | 28.67 |
| Fingerprint | 26.84 | 26.72 | 26.85 |
| Average | 28.4 | 28.19 | 27.93 |

Table 2 shows the PSNR values on the set 12 for different noise levels for different noise levels (sigma = 15, 25, 50); it can be noticed that the training method suffers from strong noise level during training, this is mainly due to the block-matching mismatch that occurs because of high noise levels, moderate noise levels like 15 and 25, produces a more stability during the training. Though the primary goal of this work is to explore no ground truth training method, the final aim is to attenuate noise perturbation in images. For reference, we compare the denoising performance our method with two non-local self-similarity based denoising methods, BM3D [14] and Non-Local Means NLM [13], the comparison gives us a fair placement of the denoising performance. Our method attenuation performs better than a classic NLM. Figures 5, 6 and 7 show further visual comparison with BM3D denoising results.

*Table 2. Comparison of different NSS based methods on Set12, Result are given in terms of PSNR (DB).*

| Images | Ours | NLM | BM3D | Ours | NLM | BM3D | Ours | NLM | BM3D |
|---|---|---|---|---|---|---|---|---|---|
| | Sigma = 15 | | | Sigma = 25 | | | Sigma = 50 | | |
| Cameraman | 27.42 | 30.49 | 31.92 | 26.75 | 28.25 | 29.45 | 24.35 | 24.14 | 26.62 |
| House | 33.22 | 33.25 | 34.95 | 31.95 | 30.63 | 32.85 | 28.92 | 26.05 | 29.69 |
| Peppers | 29.8 | 31.12 | 32.7 | 28.51 | 28.45 | 30.16 | 25.36 | 24.42 | 26.68 |
| Starfish | 28.57 | 29.61 | 31.15 | 27.09 | 26.82 | 28.56 | 23.91 | 22.92 | 25.04 |
| Monarch | 28.92 | 30.5 | 31.86 | 27.97 | 27.83 | 29.25 | 25.35 | 23.7 | 25.82 |
| Jetplane | 27.23 | 30.01 | 31.08 | 26.58 | 27.25 | 28.42 | 24.23 | 23.25 | 25.1 |
| Parrot | 27.75 | 30.2 | 31.38 | 26.95 | 27.87 | 28.93 | 24.31 | 24.64 | 25.9 |
| Lena | 27.13 | 32.53 | 34.27 | 30.62 | 30.07 | 32.07 | 25.22 | 26.66 | 29.05 |
| Barbara | 31.31 | 31.07 | 33.11 | 29.83 | 28.03 | 30.71 | 26.55 | 24.1 | 27.22 |
| Boat | 30.4 | 30.39 | 32.14 | 28.83 | 27.93 | 29.9 | 25.45 | 24.56 | 26.78 |
| Man | 30.36 | 30.48 | 31.93 | 28.87 | 28.05 | 29.61 | 26.28 | 25.03 | 26.81 |
| Fingerprint | 28.51 | 28.49 | 30.29 | 26.84 | 25.95 | 27.7 | 23.89 | 21.41 | 24.52 |
| Average | 29.22 | 30.68 | 32.23 | 28.4 | 28.09 | 29.8 | 25.32 | 24.24 | 26.6 |
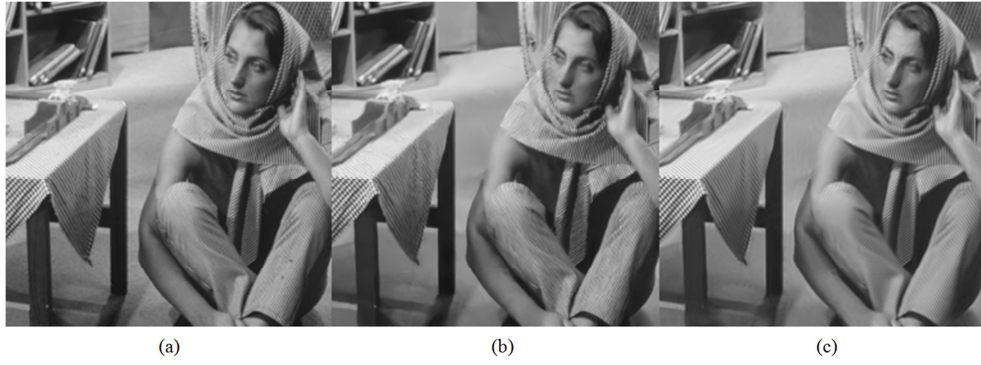
**Figure 4.** *Denoising on image Barbara with noise level 25. (a) ground truth, (b) denoised with DnCNN and (c) our method result.*



**Figure 5.** *Denoising results of the image "Parrot" with noise level 25. Left (ground truth), Middle (denoised with BM3D) and Right (our method result).*



**Figure 6.** *Denoising results of the image "House" with noise level 50. Left (ground truth), Middle (denoised with BM3D) and Right (our method result).*



**Figure 7.** *Denoising results of the image "fingerprint" with noise level 50. Left (ground truth), Middle (denoised with BM3D) and Right (our method result).*

Both Networks have been trained for 50 epochs using Stochastic Gradient Decent with momentum 0.9, weight decay 0.0001 and a 64 mini-batch. All trainings have been conducted using MatConvNet [23], under Matlab (R2018a) environment running on a PC with Intel (R) Xeon (R) CPU E5-2683 v4 @ 2.10 GHz and Nvidia GeForce GTX 1080 Ti. Trained models and codes can be found at the Github link *https://github.com/AbderraoufKhodja/SNT-NoGroundTruthTraining.git.*

## 5. Discussion

While experimenting on only single instances of noisy samples and without statistical modelling of the noise

corruption, we still managed to achieve good denoising result. In Figure 4 we can see that our method succeeds in recovering most of detail of the image Barbara, which exhibits abundant repetitive patterns. Visually, compared with state-of-the-art denoisers like DnCNN [7] we can notice that our method has a better recovery performance specially in those repetitive patterns regions, the major drawback is that the

method has more difficulties in recovering smooth regions and introduces noise artifacts which diminishes the denoising performance. The proposed framework is simple and straightforward, there is still room for improvement either in terms of a better exploitation of the non-local self-similarity prior or by introducing a statistical prior, we think that combining with [20] can benefit both frameworks, we would address these aspects in future works.

# 6. Conclusion

Image denoising is a classic problem that has been extensively studied in the scientific literature, the adoption of deep learning based method open a new door for improvement and investigation. Deep learning denoisers performs remarkably well in a supervised regime, but quickly fall short when no labeled dataset is at hand. Apart from the high restoration performance, the supervised nature of the training in most formulation of the problem makes it challenging to optimally deploy it in different scenarios. To that end we proposed a simple yet practical approach to train denoising CNN without ground truth data and without statistical modeling of the noise corruption, given only single instances of noisy images. We hope that this work would further invoke inspiration in the advancement of unsupervised method for training denoising models as to better fill the gap between classic method and modern deep learning technics in terms of flexibility and practicality in various settings. For future work we would investigate ways to improve training performance and adopting more challenging types of noise.

# Acknowledgements

# References

[1]  He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. IEEE Conference on Computer Vision and Pattern Recognition 2016, 2016-Decem, pp. 770-778.

[2]  Krizhevsky, A.; Sutskever, I.; Hinton, G. E. ImageNet classification with deep convolutional neural networks. Commun. ACM 2017, 2017, 60, pp. 84–90, doi: 10.1145/3065386.

[3]  Szegedy, C.; Ioffe, S.; Vanhoucke, V.; Alemi, A. A. Inception-v4, inception-ResNet and the impact of residual connections on learning. 31st AAAI Conf. Artif. Intell. AAAI, 2017, pp. 4278-–4284.

[4]  Russakovsky, O.; Deng, J.; Su, H.; Krause, J.; Satheesh, S.; Ma, S.; Huang, Z.; Karpathy, A.; Khosla, A.; Bernstein, M.; et al. ImageNet Large Scale Visual Recognition Challenge. Int. J. Comput. Vis., 2015, 115, pp. 211–252, doi: 10.1007/s11263-015-0816-y.

[5]  Deng, J.; Dong, W.; Socher, R.; Li, L.-J.; Kai Li; Li Fei-Fei ImageNet: A large-scale hierarchical image database. IEEE. Conference on Computer Vision and Pattern Recognition., 2009, pp. 248–255.

[6]  Girshick, R.; Donahue, J.; Darrell, T.; Malik, J. Rich feature hierarchies for accurate object detection and semantic segmentation. IEEE. Conference on Computer Vision and Pattern Recognition., 2014, 580-587.

[7]  Zhang, K.; Zuo, W.; Chen, Y.; Meng, D.; Zhang, L. Beyond a Gaussian denoiser: Residual learning of deep CNN for image denoising. IEEE Trans. Image Process., 2017, 26, 3142–3155, doi: 10.1109/TIP.2017.2662206.

[8]  Chen, Y.; Pock, T. Trainable Nonlinear Reaction Diffusion: A Flexible Framework for Fast and Effective Image Restoration. IEEE Trans. Pattern Anal. Mach. Intell., 2017, 39, 1256–1272, doi: 10.1109/TPAMI.2016.2596743.

[9]  Mao, X. J.; Shen, C.; Yang, Y. Bin Image restoration using very deep convolutional encoder-decoder networks with symmetric skip connections. Advances in Neural Information Processing Systems, 2016.

[10]  Divakar, N.; Babu, R. V. Image Denoising via CNNs: An Adversarial Approach. IEEE. Conference on Computer Vision and Pattern RecognitionW., 2017, 2017-July, pp. 1076-1083.

[11]  Xie, J.; Xu, L.; Chen, E. Image denoising and inpainting with deep neural networks. Advances in Neural Information Processing Systems, 2012, 1, pp. 341–349.

[12]  Kim, J.; Lee, J. K.; Lee, K. M. Accurate image super-resolution using very deep convolutional networks. IEEE Conference on Computer Vision and Pattern Recognition, 2016, 2016-Decem pp. 1646–1654.

[13]  Buades, A.; Coll, B.; Morel, J. M. A non-local algorithm for image denoising. In Proceedings of the Proceedings - 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR 2005; IEEE Computer Society, 2005, 2, pp. 60–65.

[14]  Dabov, K.; Foi, A.; Katkovnik, V.; Egiazarian, K. Image denoising by sparse 3-D transform-domain collaborative filtering. IEEE Trans. Image Process., 2007, 16, pp. 2080–2095, doi: 10.1109/TIP.2007.901238.

[15]  Dong, W.; Zhang, L.; Shi, G.; Li, X. Nonlocally centralized sparse representation for image restoration. IEEE Trans. Image Process., 2013, 22, pp. 1620–1630, doi: 10.1109/TIP.2012.2235847.

[16]  Gu, S.; Zhang, L.; Zuo, W.; Feng, X. Weighted nuclear norm minimization with application to image denoising. IEEE Conference on Computer Vision and Pattern Recognition, 2014, pp. 2862-2869.

[17]  Zoran, D.; Weiss, Y. From learning models of natural image patches to whole image restoration. International Conference on Computer Vision, 2011, 22, pp. 479-486.

[18] Chen, J.; Chen, J.; Chao, H.; Yang, M. Image Blind Denoising with Generative Adversarial Network Based Noise Modeling. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018, June 2018, pp. 3155-3164.

[19] Lehtinen, J.; Munkberg, J.; Hasselgren, J.; Laine, S.; Karras, T.; Aittala, M.; Aila, T. Noise2Noise: Learning image restoration without clean data. 35th Int. Conf. Mach. Learn. ICML 2018, 2018, 7, pp. 4620–4631.

[20] Soltanayev, S.; Chun, S. Y. Training deep learning based denoisers without ground truth data. Advances in Neural Information Processing Systems, 2018, 2018-December, pp. 3257-3267.

[21] Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. 3rd International Conference on Learning Representations, ICLR - Conference Track Proceedings, 2015.

[22] Roth, S.; Black, M. J. Fields of experts. Int. J. Comput. Vis., 2009, 82, pp. 205–229, doi: 10.1007/s11263-008-0197-6.

[23] Vedaldi, A.; Lenc, K. MatConvNet: Convolutional neural networks for MATLAB. In Proceedings of the MM 2015 - Proceedings of the 2015 ACM Multimedia Conference; Association for Computing Machinery, Inc, 2015, pp. 689–692.