One-Pot Multi-Frame Denoising

Lujia Jin[123]
jinlujia@pku.edu.cn
Shi Zhao[4]
magishe@pku.edu.cn
Lei Zhu[123]
zhulei@stu.pku.edu.cn
Qian Chen[123]
chen_qian@stu.pku.edu.cn
Yanye Lu[✉235]
yanye.lu@pku.edu.cn

[1] Department of Biomedical Engineering, Peking University, Beijing, China
[2] Institute of Medical Technology, Peking University, Beijing, China
[3] Institute of Biomedical Engineering, Peking University Shenzhen Graduate School, Shenzhen, China
[4] School of Physics, Peking University, Beijing, China
[5] National Biomedical Imaging Center, Peking University, Beijing, China

Abstract

The performance of learning-based denoising largely depends on clean supervision. However, it is difficult to obtain clean images in many scenes. On the contrary, the capture of multiple noisy frames for the same field of view is available and often natural in real life. Therefore, it is necessary to avoid the restriction of clean labels and make full use of noisy data for model training. So we propose an unsupervised learning strategy named one-pot denoising (OPD) for multi-frame images. OPD is the first proposed unsupervised multi-frame denoising (MFD) method. Different from the traditional supervision schemes including both supervised Noise2Clean (N2C) and unsupervised Noise2Noise (N2N), OPD executes mutual supervision among all of the multiple frames, which gives learning more diversity of supervision and allows models to mine deeper into the correlation among frames. N2N has also been proved to be actually a simplified case of the proposed OPD. From the perspectives of data allocation and loss function, two specific implementations, random coupling (RC) and alienation loss (AL), are respectively provided to accomplish OPD during model training. In practice, our experiments demonstrate that OPD behaves as the SOTA unsupervised denoising method and is comparable to supervised N2C methods for synthetic Gaussian and Poisson noise, and real-world optical coherence tomography (OCT) speckle noise.

1 Introduction

Due to the non-definite nature of image denoising, it is always difficult for methods based on reasoning to perform as well as expected on a lot of scenes. The turning milestone appears in the rise of deep learning, which greatly develops image denoising and meanwhile ignites the need for data. Convolutional neural networks (CNN) with various structures and characteristics have been designed [1, 14, 25, 39, 48]. They do not care about the causal inference of noise pattern but learn end-to-end from the noise image to its clean counterpart. However,

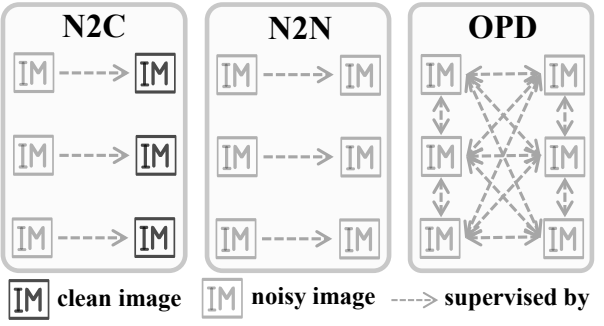

Figure 1: Comparison between our OPD and other supervision strategies, including supervised N2C and unsupervised N2N. The OPD establishes mutual supervision among multiple frames, which allows the hidden clean image to be better estimated.

clean images are difficult to obtain, making conventional learning under Noise2Clean (N2C) almost impossible to proceed. Noise2Noise (N2N)[26] overcomes this obstacle at the cost of one more noisy image. The noisy-clean image pair is replaced by the two paired noisy images to train the model under the N2N strategy.

Although it is hard to get a clean image in real-life scenes, in many cases the acquisition of multi-frame noisy images is available and even natural, such as exposure bracketing[3], astrophotography[13], and optical coherence tomography (OCT)[34], etc. Denoising for these scenes is called multi-frame denoising (MFD), which aims at finding a mapping $\boldsymbol{f}$ with a given multi-frame dataset $\left\{\left(\boldsymbol{x}_i+\boldsymbol{n}_i^j,\boldsymbol{y}_i\right)|i\in[1,N],j\in[1,m]\right\}$ such that $\boldsymbol{f}\left[\bigcup_{j\in[1,m]}\left(\boldsymbol{x}_i+\boldsymbol{n}_i^j\right)\right]=\boldsymbol{y}_i, i\in[1,N]$. As can be seen from the above problem statement, MFD is essentially a reasonable fusion of multiple noisy frames corresponding to the same underlying clean reference. All previous MFD methods[29, 31, 44] are supervised methods. For the first time, we propose an unsupervised strategy, one-pot denoising (OPD), to achieve high-quality denoising of multiple frames. As demonstrated in Fig. 1, OPD strives to extend the supervision from noisy-clean pairs to a group of all noisy frames corresponding to the same clean target. Specifically, we design a data allocation method that continuously shuffles the supervision pair from multiple frames during training so that the information contained in each frame could be fully utilized. This implementation is named OPD-random coupling (OPD-RC). In addition, we construct an alienation loss function, which has an equivalent denoising effect to OPD-RC and provide a more formulaic and therefore more intuitive understanding for the OPD strategy. This implementation is named OPD-alienation loss (OPD-AL). Both implementations of OPD are experimented on additive white Gaussian noise (AWGN), signal-dependent Poisson noise, and OCT speckle noise. Experimental results on all three noises show that, both qualitatively and quantitatively, OPD performs better than N2N and sometimes outperforms N2C.

In a nutshell, the main contributions of our work are as follows:

1. We propose OPD, a denoising strategy based on an unprecedented mutual supervision paradigm. OPD is the first proposed unsupervised MFD method.
2. From the perspectives of data allocation and loss function, two specific implementations, OPD-RC and OPD-AL are presented for MFD. We also reveal that the well-known N2N[26] can be interpreted as a simplified case of our OPD.

3. Experiments show that our OPD behaves as the SOTA unsupervised denoising method and is comparable to supervised N2C methods for several MFD tasks, including denoising AWGN and Poisson noise, and OCT speckle noise reduction.

2 Related Works

2.1 Multi-frame Denoising

Compared to single-image denoising (SID), MFD has received significantly less attention in the past decade. Tico[40] first migrated the popular Non-Local Means (NLM)[5] from single image denoising to MFD. This method compares the similarity of blocks not only within but also among frames. V-BM3D[22] and V-BM4D[27, 28] are based on the famous BM3D[9] to denoise videos through sparse 3D transform-domain collaborative filtering. Buades et al. [7] provided a complex processing chain including accurate registration and noise estimation. Hasinoff et al.[16] applied an FFT-based alignment algorithm and a hybrid 2D/3D Wiener filter to burst denoising, which had been built atop Android's Camera2 API. Completely different from the aforementioned block-based fusion methods, accumulation after registration (AAR)[6] directly uses weighted averaging to fuse multiple frames, which is proved to be effective in zero-mean noise reduction.

All of the above are non-learning methods, while learning-based methods are rarer. Godard et al.[12] constructed a simple but effective recurrent neural network inspired by series data processing. Mildenhall et al.[31] proposed a kernel prediction network (KPN) to produce clean images from bursts by unique 3D denoising kernels. Marinc et al.[29] used multi-scale kernels to extend KPN to multi-KPN (MKPN). Furthermore, Xia et al.[44] developed a basis prediction network (BPN) for effective burst denoising with large kernels, which achieves both significant quality improvement and faster run-time.

2.2 Supervision Strategies for Image Denoising

Most learning-based image denoising methods follow the conventional N2C supervision paradigm[1, 14, 25, 39, 48], which makes it indispensable to obtain clean images as labels. N2N[26] breaks this limitation by exploring an alternative supervision paradigm, in which pairs of noisy images corresponding to a common unknown clean target are used for training. Wu et al.[43] showed that the results of optimization are equivalent under N2N and N2C as long as the amount of data is large enough. AltN2N[8] improves the performance of N2N under limited data by fine tuning. In the past two years, N2N has been widely used in low dose computed tomography[15], positron emmision tomography[19], synthetic aperture rader[10] and other image denoising tasks. Speech denoising[20], video enhancement[4, 46], nanochennel measurement[38] and other non-image denoising has also excellently proceeded under N2N.

Going further than N2N, self-supervised strategies denoise with only a single noisy image, and they can be divided into two categories. The first category of methods utilizes priori noise models as an external aid to construct another noisy image and pairs it with the provided data for further N2N learning. Noiser2Noise[32] and Noise-as-Clean[45] are representative methods in this category. These methods are strictly limited by prior noise knowledge. The second category of methods does not require any additional knowledge. Methods based on mask prediction, such as DIP[41], Noise2Void (N2V)[23], Noise2Self

(N2S)[2] and blind-splot network[24], are typical representatives of this category, but their denoising performance is inferior to the aforementioned prior knowledge-assisted methods.

3 Methods

In this Section, we first retrospect and formulate three conventional strategies for MFD as preliminaries for subsequent method development (Sec. 3.1). Then the principle of our OPD and two specific implementations, OPD-RC and OPD-AL, are introduced (Sec. 3.2). Finally, we compare OPD with other strategies to reasonably analyze their pros and cons (Sec. 3.3).

3.1 Retrospecting Conventional MFD

MFD aims to learn the underlying clean target $\boldsymbol{x}_i$ based on the multiple noisy images $\mathbb{X}: \left\{\boldsymbol{x}_i+\boldsymbol{n}_i^j \mid i\in[1,N],\ j\in[1,m]\right\}$, where N denotes the number of samples and m refers to the number of noisy frames per sample.

From the perspective of supervision strategies, there are mainly three existing strategies for MFD: AAR[6], N2C and N2N[26], which are described in detail as follows.

AAR for MFD: As one of the most classical methods, AAR[6] is a simple but effective algorithm. After registering all frames corresponding to the same sample, a clean $\boldsymbol{x}_i$ can be estimated simply by accumulating and averaging as:

$$\widehat{\boldsymbol{x}}_i=\frac{1}{m}\sum_{j=1}^{m}(\boldsymbol{x}_i+\boldsymbol{n}_i^j),\quad i=1,2,...,N \tag{1}$$

N2C for MFD: Learning-based methods improve the generalization potential. Based on the clean estimation of AAR, a set of training pairs $\mathbb{T}_{N2C}: \left\{(\boldsymbol{x}_i+\boldsymbol{n}_i^j,\ \widehat{\boldsymbol{x}}_i) \mid \ i\in[1,N],\ j\in[1,m]\right\}$ can be built and N2C learning can be performed with the loss:

$$\mathcal{L}_{N2C}=\frac{1}{N\times m}\sum_{i=1}^{N}\sum_{j=1}^{m}\left\|f_{\boldsymbol{\Theta}}(\boldsymbol{x}_i+\boldsymbol{n}_i^j)-\widehat{\boldsymbol{x}}_i\right\|_2^2, \tag{2}$$

where L_2 error is used by default in our derivation.

N2N for MFD: Among the strategies that help models escape the constraints of clean supervision, N2N is the most representative. Find a random permutation $\mathbb{I}_i$ of the sequence $\mathbb{I}=[1,2,...,m]$ for each $i\in[1,N]$. $\mathbb{J}_i$ and $\mathbb{K}_i$ are two sequences obtained by equally dividing $\mathbb{I}_i$, which means that $\mathbb{J}_i$ and $\mathbb{K}_i$ constitute a random uniform partition of $\mathbb{I}$. Note that when m is odd, randomly discard an element in $\mathbb{I}$. Treat the elements in $\mathbb{J}_i$ and $\mathbb{K}_i$, $i\in[1,N]$ as indexes of frames and equally divide the multi-frame data into two parts:

$$\begin{aligned}\mathbb{X}_1&:\left\{\boldsymbol{x}_i+\boldsymbol{n}_i^j \mid i\in[1,N],\ j\in\mathbb{J}_i\right\}\\ \mathbb{X}_2&:\left\{\boldsymbol{x}_i+\boldsymbol{n}_i^k \mid i\in[1,N],\ k\in\mathbb{K}_i\right\}\end{aligned} \tag{3}$$

where it should be noted that j and k are just the elements in $\mathbb{J}_i$ and $\mathbb{K}_i$ but not the indexes.

Then the elements in $\mathbb{X}_1$ and $\mathbb{X}_2$ can be paired one-to-one to construct the training set $\mathbb{T}_{N2N}$ and N2N learning can be performed with the loss:

$$\mathcal{L}_{N2N} = \frac{1}{N \times \lfloor \frac{m}{2} \rfloor} \sum_{i=1}^{N} \sum_{\substack{j \in \mathbb{J}_i \\ k \in \mathbb{K}_i}} \left\| f_{\mathbf{\Theta}}(\boldsymbol{x}_i + \boldsymbol{n}_i^j) - (\boldsymbol{x}_i + \boldsymbol{n}_i^k) \right\|_2^2, \quad s.t. \quad idx(j) = idx(k), \tag{4}$$

where $idx(j)$ and $idx(k)$ represent the corresponding indexes of j and k in $\mathbb{J}_i$ and $\mathbb{K}_i$.

3.2 One-Pot Multi-Frame Denoising

In Eq. (4), two ways can be found to use a pair of noisy images, which are employing one to supervise the other and vice versa. Based on this consideration, we propose a concept of "mutual supervision". As shown by the bidirectional arrow in Fig. 1, the roles of the two noisy images participating in the training under mutual supervision are not absolutely prescribed, but interchanged, entangled and equivalent. Furthermore, for multi-frame scenarios with $m > 2$, mutual supervision can be established among all noisy images corresponding to the same $\boldsymbol{x}_i$. The learning strategy based on the above concept is evocatively named OPD. OPD enables each noisy image to play an equally important role. Diversified samples and labels enable the model to squeeze out much more hidden inter-frame information contained in the data during learning. At the same time, this also makes the model face more optimization possibilities, which potentially leads to an improvement in denoising performance.

The most intuitive way to perform OPD is to go through all the one-to-one unidirectional supervision pairs. Nevertheless, it is easy to realize that skyrocketing data size makes this way so crude. Considering that the learnable models usually look for the minimum in an iterative manner on the hypersurface defined by the loss function, from the perspectives of reconstruction of the iterative data pairs and reconstruction of the loss function, we propose two feasible OPD implementation methods, OPD-RC and OPD-AL, respectively.

OPD-RC: Since the data is fed into the model iteratively during training, we can simply shuffle the multiple frames each time before a new iteration to continuously reconstruct the supervision direction. Assuming that s refers to a step during model updating, before the sth iteration, construct random and uniform partition $\mathbb{J}_i^s$ and $\mathbb{K}_i^s$ of the sequence $\mathbb{I} = [1, 2, ..., m]$. Then randomly divide the m noisy frames corresponding to a same $\boldsymbol{x}_i$ into two sets:

$$\begin{aligned} \mathbb{X}_1^s &: \left\{ \boldsymbol{x}_i + \boldsymbol{n}_i^j \middle| i \in [1, N], j \in \mathbb{J}_i^s, s \in \mathbb{N}^* \right\} \\ \mathbb{X}_2^s &: \left\{ \boldsymbol{x}_i + \boldsymbol{n}_i^k \middle| i \in [1, N], k \in \mathbb{K}_i^s, s \in \mathbb{N}^* \right\} \end{aligned} \tag{5}$$

According to the back-propagation rule, the model at step $s+1$ can be updated as:

$$\mathbf{\Theta}^{s+1} = \mathbf{\Theta}^s - \eta \frac{\partial}{\partial \mathbf{\Theta}} \left[\sum_{i=1}^{N_B} \sum_{\substack{j \in \mathbb{J}_i^s \\ k \in \mathbb{K}_i^s}} \left\| f_{\mathbf{\Theta}}(\boldsymbol{x}_i + \boldsymbol{n}_i^j) - (\boldsymbol{x}_i + \boldsymbol{n}_i^k) \right\|_2^2 \right], \quad s.t. \quad idx(j) = idx(k), \tag{6}$$

where $\mathbf{\Theta}^s$ is the parameters at the sth step, η means the learning rate and N_B is the batch size. j_l^s and k_l^s are the indexes of the input and the label randomly coupled before the sth step.

OPD-RC makes each of the m noisy frames appear in each iteration with equal probability and serve as input or label with the same chance. This operation greatly extends the diversity of data pairing and supervision without any more training time consumption. As

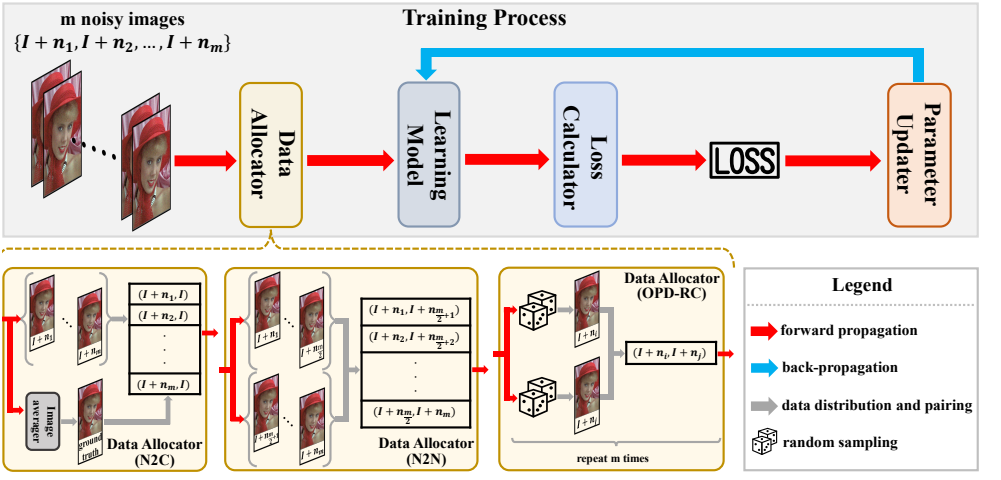

Figure 2: The workflows of the proposed OPD-RC and other denoising strategies including N2C and N2N. The upper part presents the common workflow to train a learning-based model. The data allocator represented by the gold box is the key difference between OPD-RC and other strategies, which are shown in detail in the three small sub-figures at the bottom.

long as the training process goes through enough iterations, it is reasonable to think that the multi-frame images are evenly used and the mutual supervision among them has been established in a practical sense. Furthermore, OPD-RC does not affect the choice and design of the network architecture and loss function. Fig. 2 shows the general principle of OPD-RC and its comparison with N2C and N2N. The data pairing way in the data allocator is the key difference between OPD-RC and other supervision strategies, as shown in the lower part of Fig. 2.

OPD-AL: In order to make the m noisy frames corresponding to an $\boldsymbol{x}_i$ play a full and balanced role during training, the second averaging operation in Eq. (2) can be moved to $f_{\boldsymbol{\Theta}}(\cdot)$:

$$\mathcal{L}^C_{OPD} = \frac{1}{N_B}\sum_{i=1}^{N_B}\left\|\frac{1}{m}\sum_{j=1}^{m} f_{\boldsymbol{\Theta}}(\boldsymbol{x}_i+\boldsymbol{n}_i^j)-\widehat{\boldsymbol{x}}_i\right\|_2^2, \tag{7}$$

where the superscript C of $\mathcal{L}^C_{OPD}$ indicates that this loss is still under clean supervision.

According to the polynomial theorem, reorganize $\mathcal{L}^C_{OPD}$:

$$\mathcal{L}^C_{OPD} = \frac{1}{N_B}\sum_{i=1}^{N_B}\left[\frac{1}{m}\sum_{j=1}^{m}\left\|\boldsymbol{y}_i^j-\widehat{\boldsymbol{x}}_i\right\|_2^2-\frac{1}{m^2}\sum_{j=1}^{m-1}\sum_{k=j+1}^{m}\left\|\boldsymbol{y}_i^j-\boldsymbol{y}_i^k\right\|_2^2\right], \tag{8}$$

where $f_{\boldsymbol{\Theta}}(\boldsymbol{x}_i+\boldsymbol{n}_i^j)$ and $f_{\boldsymbol{\Theta}}(\boldsymbol{x}_i+\boldsymbol{n}_i^k)$ are replaced by $\boldsymbol{y}_i^j$ and $\boldsymbol{y}_i^k$ in writing, respectively. The specific derivation from Eq. (7) to Eq. (8) is provided in Supplementary Material (S.M).

Use mean square error (MSE) and mean square alienation (MSA) to replace the two items included in the first summation in Eq. (8):

$$\mathcal{L}^C_{OPD} = \frac{1}{N_B}\sum_{i=1}^{N_B}\left[(\mathcal{L}^C_{MSE})_i-(\mathcal{L}_{MSA})_i\right] \tag{9}$$

Since Wu et al.[43] have proved the equivalence of convergence with clean or noisy labels, replace $\widehat{\boldsymbol{x}}_i$ with its corresponding noisy frames, we can finally get the OPD loss:

$$\mathcal{L}_{OPD} = \frac{1}{N_B}\sum_{i=1}^{N_B}\left[(\mathcal{L}_{MSE}^{N})_i - (\mathcal{L}_{MSA})_i\right], \tag{10}$$

where $(\mathcal{L}_{MSE}^{N})_i$ and $(\mathcal{L}_{MSA})_i$ are respectively formulated as:

$$\begin{aligned}(\mathcal{L}_{MSE}^{N})_i &= \frac{1}{m(m-1)}\sum_{j=1}^{m}\sum_{\substack{k=1,\\ k\neq j}}^{m}\left\|\boldsymbol{y}_i^j - (\boldsymbol{x}_i + \boldsymbol{n}_i^k)\right\|_2^2 \\ (\mathcal{L}_{MSA})_i &= \frac{1}{m^2}\sum_{j=1}^{m-1}\sum_{k=j+1}^{m}\left\|\boldsymbol{y}_i^j - \boldsymbol{y}_i^k\right\|_2^2\end{aligned} \tag{11}$$

So far, OPD loss has been successfully constructed. The key constraint on inter-frame mutual supervision is the $\mathcal{L}_{MSA}$ term, which rewards the inter-frame alienation mined by the model. When m equals to 2, Eq. (11) is reduced to the loss of N2N superimposed with mutual supervision. This shows that the proposed OPD is a generalized form of N2N.

3.3 OPD vs. other Supervision Strategies

OPD vs. AAR: As a learning-based strategy, OPD has no restrictions on the physical pattern of noise. However, AAR can only denoise images with zero-mean signal-independent noise. Compared with OPD, AAR is not suitable for generalization, since numerous frames under the same view are required when denoising on new data.

OPD vs. N2C: As an unsupervised strategy, OPD does not require any clean images as labels, which is not the case for N2C. In addition, N2C regards m noisy frames corresponding to the same clean target as independent samples, whereas OPD regards them as a whole for more global consideration and more comprehensive mining.

OPD vs. N2N: Compared with OPD, N2N does not fully utilize multi-frame data. Pairwise matching in m frames and roles for input and label are both determined arbitrarily before training. It is easy to realize that the m noisy images corresponding to each $\boldsymbol{x}_i$ are equally valuable and should play an equal role, which is exactly what OPD achieves.

4 Experiments

Our OPD is experimented in three typical scenarios: synthetic Gaussian and Poisson noise, and OCT speckle noise. Representatives of non-learning (e.g. NLM[5], BM3D[9]), supervised (e.g. N2C, KPN[31]) and unsupervised (e.g. N2N[26], N2S[2]) denoising methods participate in the comparison, including single-image and multi-frame algorithms. All the quantitative evaluation results in this paper are statistically significant. More results are presented in S.M, and high-resolution versions of Fig. 3 and Fig. 4 are also provided in S.M.

4.1 Settings

Datasets: For synthetic noise, the clean data comes from 50,000 images in the ImageNet [36] validation set, which are cropped into 256×256. We randomly add AWGN with $\sigma = 25$ or Poisson noise with $\lambda = 30$ to the images and the frame number is set to 8. The same process is implemented on BSD300[30], KODAK[1] and SET14[47] to build testing sets.

[1] http://r0k.us/graphics/kodak/

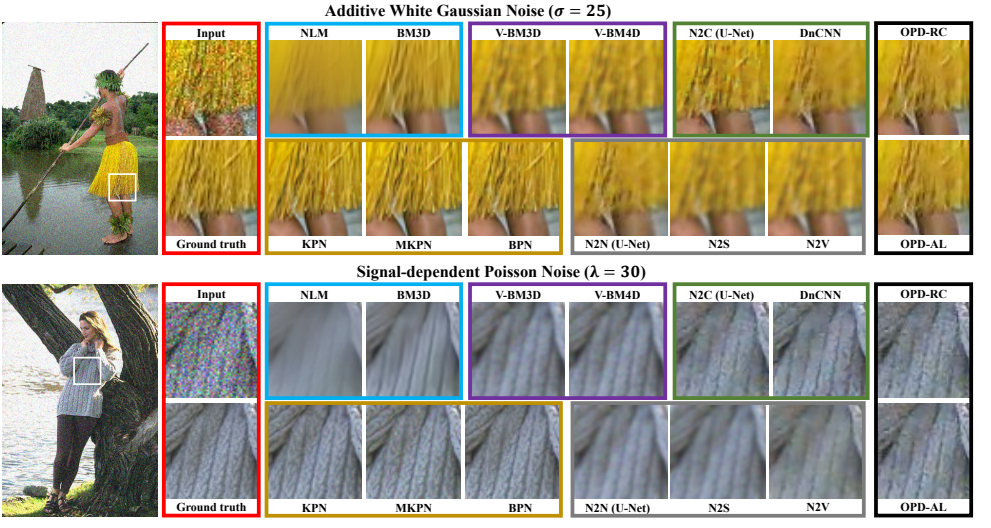

Figure 3: Example results of denoising synthetic noise. The results of different categories of methods are framed with different colored boxes. The categories are listed in Tab. 1.

For OCT speckle noise reduction, the data comes from PKU37[11]. Due to the inconsistent number of frames, we only use 16 frames per sample to align frame numbers among different samples, which means our experiments use only one-third the amount of PKU37.

Implementation Details: Considering the training efficiency, a modified U-Net[18, 35] was chosen to be the demonstrative model for N2C, N2N, OPD-RC and OPD-AL (See S.M for network architecture). He's method[17] was used for initialization. Adam[21] was used for parameter optimization with L_2 loss. In all experiments, one-tenth of the data was randomly split from the training set to be validation set. All our experiments were conducted based on PyTorch[33]. Six NVIDIA RTX 3090 graphical cards each with 24GB memory were used. The hyperparameters for each experiment were different, which could be found in S.M. Peak signal-to-noise ratio (PSNR), structural similarity (SSIM)[42], and root-mean-square error (RMSE) were used as evaluation metrics to quantify the performance of involved methods.

4.2 Denoising Synthetic Noise

The average quantitative evaluation for the three testing sets with either Gaussian or Poisson noise are shown in Tab. 1. Fig. 3 shows example results. Numerous representative methods are involved in the comparison. According to the supervision scheme, they are divided into non-learning, supervised and unsupervised methods. According to the image scene, they are divided into single-image methods and multi-frame methods. Next, we compare and analyze OPD and other methods of various categories.

OPD vs. other Unsupervised Methods: Both quantitative and qualitative results show that the proposed OPD achieves SOTA among all unsupervised methods. As can be seen from Tab. 1, for Gaussian noise, compared with N2N, the PSNR of OPD-AL is improved by 0.88dB, the SSIM is improved by 0.020, and the RMSE is decreased by 0.009. Similar boosts also occur on OPD-RC and for Poisson noise. Small changes in metrics show big changes in visual perception. Fig. 3 shows that both OPD algorithms preserve high-frequency details

Table 1: The quantitative evaluation results of denoising synthetic noise and OCT speckle noise. For each scenario, the globally highest and second highest results are denoted as **red** and **blue**, respectively. Locally for unsupervised methods, the highest and the second highest results are labeled with **double underline** and **wave underline**, respectively.

Category		Method	Gaussian			Poisson			OCT		
			PSNR	SSIM	RMSE	PSNR	SSIM	RMSE	PSNR	SSIM	RMSE
Input			22.72	0.505	0.074	21.20	0.469	0.088	20.35	0.513	0.096
non-learning	single-image	NLM[5]	24.92	0.670	0.059	24.96	0.676	0.058	26.36	0.600	0.048
		BM3D[9]	25.63	0.774	0.055	23.82	0.684	0.066	26.67	0.612	0.047
	multi-frame	V-BM3D[22]	27.50	0.801	0.051	25.56	0.707	0.062	27.62	0.623	0.044
		V-BM4D[27]	27.86	0.811	0.051	25.79	0.711	0.062	27.87	0.630	0.043
supervised	single-image	N2C	28.04	0.798	0.041	27.92	0.781	0.041	29.79	0.898	0.033
		DnCNN[48]	29.01	0.827	0.036	28.39	0.814	0.039	28.84	**0.871**	0.036
	multi-frame	KPN[31]	32.31	0.917	0.025	32.28	0.916	**0.025**	26.68	0.582	0.047
		MKPN[29]	**32.67**	**0.924**	**0.024**	**32.43**	**0.923**	**0.025**	28.68	0.592	0.037
		BPN[44]	**33.84**	**0.942**	**0.021**	**33.11**	**0.936**	**0.023**	29.00	0.602	0.036
unsupervised	single-image	N2N[26]	27.48	0.787	0.048	27.28	0.775	0.044	28.07	0.817	0.040
		N2S[2]	26.88	0.780	0.049	27.11	0.760	0.045	22.23	0.523	0.089
		N2V[23]	26.29	0.772	0.050	26.95	0.721	0.046	21.90	0.518	0.091
	multi-frame	OPD-RC	**28.15**	**0.805**	**0.040**	**28.22**	**0.789**	**0.040**	**30.69**	**0.900**	**0.029**
		OPD-AL	**28.36**	**0.807**	**0.039**	**28.16**	**0.790**	**0.040**	**30.40**	**0.871**	**0.030**

better for both Gaussian and Poisson noise, such as the fringes on the hula skirt in the upper example of Fig. 3 and the texture of the sweater in the lower example of Fig. 3.

OPD vs. Supervised Methods: It is unfair to compare unsupervised OPD with supervised methods, but we still do some discussion in order to evaluate OPD more comprehensively. Tab. 1 shows that OPD is better than N2C but worse than all other supervised methods. However, from the visual perception in Fig. 3, the denoising effect of OPD is comparable to that of N2C and DnCNN[48]. As an unsupervised method, this result is already satisfactory.

OPD vs. Non-learning Methods: Both the quantization results given in Tab. 1 and the examples shown in Fig. 3 show that OPD exhibits unquestionable denoising advantages over non-learning methods. Looking at the example in Fig. 3, it is easy to see that non-learning methods tend to suffer from oversmoothing, which is well overcome by OPD.

OPD vs. other Multi-frame Methods: OPD is the first unsupervised MFD method. Of course, it is reasonable that OPD as an unsupervised method is inferior to supervised multi-frame methods. However, real-life multi-frame scenes often do not support obtaining clean labels, which is precisely the significance of our research.

OPD vs. single-image Methods: Regardless of supervised or unsupervised, multi-frame methods always significantly outperform single-image methods. Specifically, OPD is better than N2N[26], N2S[2] and N2V[23] on detail retention. Even more astonishing, supervised multi-frame methods such as BPN[44] are almost indistinguishable from ground truth.

4.3 OCT Speckle Noise Reduction

The quantitative evaluation of OCT are shown in Tab. 1. Fig. 4 shows an example result.

OPD Wins SOTA on OCT: Among all the methods involved in the comparison, including supervised methods, OPD-RC captures the SOTA result, and OPD-AL is only marginally behind. The two OPD methods are the only methods with PSNR exceeding 30 dB. In addition, the SSIM of OPD-RC reaches 0.9 and the RMSE reaches below 0.03. Fig. 4 demonstrates

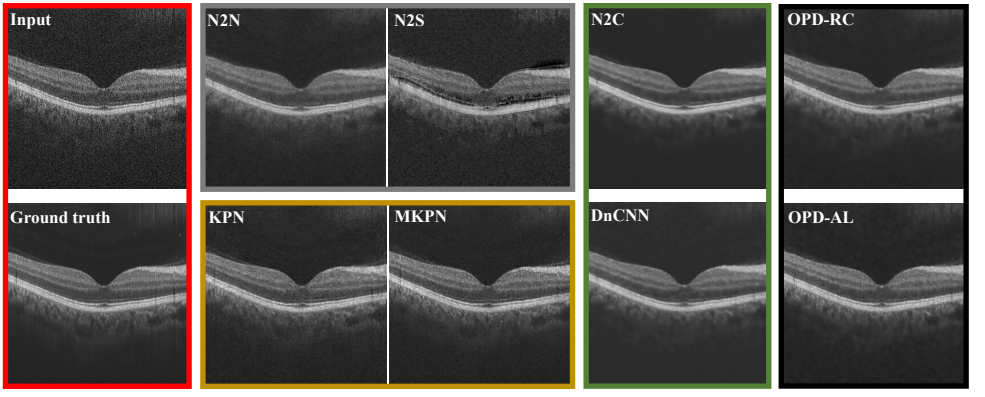

Figure 4: Example result of denoising OCT speckle noise. The results of different categories of methods are framed with different colored boxes. The categories are listed in Tab. 1.

that the OPD-denoised image has sharper and more intact retinal layers, and the complex signal of the choroid is not over-smoothed as in other methods.

Kernel-based Methods Fail: Unlike what is seen on synthetic noise, supervised multi-frame methods such as KPN[31] perform poorly on OCT. Destructive streaks appear in the images corresponding to KPN[31], MKPN[29] and BPN[44] in Fig. 4. This is because the laser coherent noise contained in OCT is long-range[37], and methods based on local kernel prediction cannot mine such global noise well, but OPD based on mutual supervision can.

Self-supervised Methods Fail: N2S[2] and N2V[23] hardly converge on the OCT denoising task because their premise for local mask estimation is that the noise contained in the image is signal-independent, which is the opposite of the case of OCT.

OPD-RC vs. OPD-AL: Quantitatively, OPD-AL outperforms OPD-RC on synthetic noise, but vice versa on OCT. Qualitatively, OPD-AL-processed images seem to be sharper than those OPD-RC-processed, both on synthetic noise and OCT. This reflects that the alienation loss may be slightly better than simple randomization. In addition, the results also illustrate the importance of visual perception beyond quantitative evaluation.

5 Conclusion

For the first time, our work defines the concept of mutual supervision and proposes an unsupervised strategy named OPD for MFD. Unlike pairwise supervision in traditional learning strategies, OPD uniformly establishes supervision relationships among multiple images participating in learning. We propose two specific algorithms, OPD-RC and OPD-AL, respectively from the perspectives of data allocation and alienation loss design. The experiments show the effectiveness of our OPD strategy on several MFD tasks including denoising Gaussian and Poisson noise and OCT speckle noise reduction.

Acknowledgments

This work was supported in part by the Beijing Natural Science Foundation (Z210008) and in part by the Shenzhen Science and Technology Program (KQTD20180412181221912, JCYJ20200109140603831).

References

[1] Saeed Anwar and Nick Barnes. Real image denoising with feature attention. In IEEE/CVF International Conference on Computer Vision (ICCV), pages 3155–3164, 2019.

[2] Joshua Batson and Loic Royer. Noise2self: Blind denoising by self-supervision. In International Conference on Machine Learning (ICML), pages 524–533. PMLR, 2019.

[3] Marcelo Bertalmío and Stacey Levine. Fusion of bracketing pictures. In IEEE Conference for Visual Media Production (CVMP), pages 25–34. IEEE, 2009.

[4] AA Boiko and RO Malashin. Single-frame noise2noise: method of training a neural network without using reference data for video sequence image enhancement. Journal of Optical Technology, 87(10):567–573, 2020.

[5] Antoni Buades, Bartomeu Coll, and J-M Morel. A non-local algorithm for image denoising. In IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), volume 2, pages 60–65. IEEE, 2005.

[6] Antoni Buades, Yifei Lou, Jean-Michel Morel, and Zhongwei Tang. Multi image noise estimation and denoising. HAL preprint hal-00510866, 2010.

[7] Toni Buades, Yifei Lou, Jean-Michel Morel, and Zhongwei Tang. A note on multi-image denoising. In IEEE 2009 International Workshop on Local and Non-Local Approximation in Image Processing, pages 1–15. IEEE, 2009.

[8] Adria Font Calvarons. Improved noise2noise denoising with limited data. In IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pages 796–805, 2021.

[9] Kostadin Dabov, Alessandro Foi, Vladimir Katkovnik, and Karen Egiazarian. Image denoising by sparse 3-d transform-domain collaborative filtering. IEEE Transactions on Image Processing (TIP), 16(8):2080–2095, 2007.

[10] Emanuele Dalsasso, Loïc Denis, and Florence Tupin. Sar2sar: a semi-supervised despeckling algorithm for sar images. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 14:4321–4329, 2021.

[11] Mufeng Geng, Xiangxi Meng, Lei Zhu, Zhe Jiang, Mengdi Gao, Zhiyu Huang, Bin Qiu, Yicheng Hu, Yibao Zhang, Qiushi Ren, et al. Triplet cross-fusion learning for unpaired image denoising in optical coherence tomography. IEEE Transactions on Medical Imaging, 2022.

[12] Clément Godard, Kevin Matzen, and Matt Uyttendaele. Deep burst denoising. In European Conference on Computer Vision (ECCV), pages 538–554. Springer, 2018.

[13] Claire Guilloteau, Thomas Oberlin, Olivier Berné, and Nicolas Dobigeon. Hyperspectral and multispectral image fusion under spectrally varying spatial blurs–application to high dimensional infrared astronomical imaging. IEEE Transactions on Computational Imaging (TCI), 6:1362–1374, 2020.

[14] Shi Guo, Zifei Yan, Kai Zhang, Wangmeng Zuo, and Lei Zhang. Toward convolutional blind denoising of real photographs. In IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pages 1712–1722, 2019.

[15] Ahmed M Hasan, Mohammad Reza Mohebbian, Khan A Wahid, and Paul Babyn. Hybrid-collaborative noise2noise denoiser for low-dose ct images. IEEE Transactions on Radiation and Plasma Medical Sciences, 5(2):235–244, 2020.

[16] Samuel W Hasinoff, Dillon Sharlet, Ryan Geiss, Andrew Adams, Jonathan T Barron, Florian Kainz, Jiawen Chen, and Marc Levoy. Burst photography for high dynamic range and low-light imaging on mobile cameras. ACM Transactions on Graphics (TOG), 35(6):1–12, 2016.

[17] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In IEEE/CVF International Conference on Computer Vision (ICCV), pages 1026–1034, 2015.

[18] Zhe Jiang, Zhiyu Huang, Bin Qiu, Xiangxi Meng, Yunfei You, Xi Liu, Gangjun Liu, Chuangqing Zhou, Kun Yang, Andreas Maier, et al. Comparative study of deep learning models for optical coherence tomography angiography. Biomedical Optics Express, 11 (3):1580–1597, 2020.

[19] Seung-Kwan Kang, Si-Young Yie, and Jae-Sung Lee. Noise2noise improved by trainable wavelet coefficients for pet denoising. Electronics, 10(13):1529, 2021.

[20] Madhav Mahesh Kashyap, Anuj Tambwekar, Krishnamoorthy Manohara, and S Natarajan. Speech denoising without clean training data: a noise2noise approach. arXiv preprint arXiv:2104.03838, 2021.

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

[22] Dabov Kostadin, Foi Alessandro, and Egiazarian Karen. Video denoising by sparse 3d transform-domain collaborative filtering. In IEEE 15th European Signal Processing Conference (ESPC), volume 149, page 2. IEEE, 2007.

[23] Alexander Krull, Tim-Oliver Buchholz, and Florian Jug. Noise2void-learning denoising from single noisy images. In IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pages 2129–2137, 2019.

[24] Samuli Laine, Tero Karras, Jaakko Lehtinen, and Timo Aila. High-quality self-supervised deep image denoising. Advances in Neural Information Processing Systems, 32:6970–6980, 2019.

[25] Stamatios Lefkimmiatis. Universal denoising networks: a novel cnn architecture for image denoising. In IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pages 3204–3213, 2018.

[26] Jaakko Lehtinen, Jacob Munkberg, Jon Hasselgren, Samuli Laine, Tero Karras, Miika Aittala, and Timo Aila. Noise2noise: Learning image restoration without clean data. arXiv preprint arXiv:1803.04189, 2018.

[27] Matteo Maggioni, Giacomo Boracchi, Alessandro Foi, and Karen Egiazarian. Video denoising using separable 4d nonlocal spatiotemporal transforms. In Image Processing: Algorithms and Systems IX, volume 7870, page 787003. International Society for Optics and Photonics, 2011.

[28] Matteo Maggioni, Giacomo Boracchi, Alessandro Foi, and Karen Egiazarian. Video denoising, deblocking, and enhancement through separable 4-d nonlocal spatiotemporal transforms. IEEE Transactions on Image Processing (TIP), 21(9):3952–3966, 2012.

[29] Talmaj Marinč, Vignesh Srinivasan, Serhan Gül, Cornelius Hellge, and Wojciech Samek. Multi-kernel prediction networks for denoising of burst images. In IEEE International Conference on Image Processing (ICIP), pages 2404–2408. IEEE, 2019.

[30] David Martin, Charless Fowlkes, Doron Tal, and Jitendra Malik. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In IEEE/CVF International Conference on Computer Vision (ICCV), volume 2, pages 416–423. IEEE, 2001.

[31] Ben Mildenhall, Jonathan T Barron, Jiawen Chen, Dillon Sharlet, Ren Ng, and Robert Carroll. Burst denoising with kernel prediction networks. In IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pages 2502–2510, 2018.

[32] Nick Moran, Dan Schmidt, Yu Zhong, and Patrick Coady. Noisier2noise: Learning to denoise from unpaired noisy data. In IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pages 12064–12072, 2020.

[33] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, et al. Pytorch: An imperative style, high-performance deep learning library. Advances in Neural Information Processing Systems, 32:8026–8037, 2019.

[34] Bin Qiu, Zhiyu Huang, Xi Liu, Xiangxi Meng, Yunfei You, Gangjun Liu, Kun Yang, Andreas Maier, Qiushi Ren, and Yanye Lu. Noise reduction in optical coherence tomography images using a deep neural network with perceptually-sensitive loss function. Biomedical Optics Express, 11(2):817–830, 2020.

[35] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI), pages 234–241. Springer, 2015.

[36] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, et al. Imagenet large scale visual recognition challenge. International Journal of Computer Vision (IJCV), 115(3):211–252, 2015.

[37] Joseph M Schmitt, SH Xiang, and Kin Man Yung. Speckle in optical coherence tomography. Journal of Biomedical Optics, 4(1):95–105, 1999.

[38] Takayuki Takaai and Makusu Tsutsui. Unsupervised noise reduction for nanochannel measurement using noise2noise deep learning. In Pacific-Asia Conference on Knowledge Discovery and Data Mining, pages 44–56. Springer, 2021.

[39] Chunwei Tian, Yong Xu, Zuoyong Li, Wangmeng Zuo, Lunke Fei, and Hong Liu. Attention-guided cnn for image denoising. Neural Networks, 124:117–129, 2020.

[40] Marius Tico. Multi-frame image denoising and stabilization. In IEEE 16th European Signal Processing Conference (ESPC), pages 1–4. IEEE, 2008.

[41] Dmitry Ulyanov, Andrea Vedaldi, and Victor Lempitsky. Deep image prior. In IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pages 9446–9454, 2018.

[42] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. Image quality assessment: from error visibility to structural similarity. IEEE Transactions on Image Processing (TIP), 13(4):600–612, 2004.

[43] Dufan Wu, Kuang Gong, Kyungsang Kim, Xiang Li, and Quanzheng Li. Consensus neural network for medical imaging denoising with only noisy training samples. In International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI), pages 741–749. Springer, 2019.

[44] Zhihao Xia, Federico Perazzi, Michaël Gharbi, Kalyan Sunkavalli, and Ayan Chakrabarti. Basis prediction networks for effective burst denoising with large kernels. In IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pages 11844–11853, 2020.

[45] Jun Xu, Yuan Huang, Ming-Ming Cheng, Li Liu, Fan Zhu, Zhou Xu, and Ling Shao. Noisy-as-clean: learning self-supervised denoising from corrupted image. IEEE Transactions on Image Processing (TIP), 29:9316–9329, 2020.

[46] Martin Zach and Erich Kobler. Real-world video restoration using noise2noise. In Joint Austrian Computer Vision and Robotics Workshop 2020, pages 145–150. Verlag der Technischen Universität Graz, 2020.

[47] Roman Zeyde, Michael Elad, and Matan Protter. On single image scale-up using sparse-representations. In International Conference on Curves and Surfaces, pages 711–730. Springer, 2010.

[48] Kai Zhang, Wangmeng Zuo, Yunjin Chen, Deyu Meng, and Lei Zhang. Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising. IEEE Transactions on Image Processing (TIP), 26(7):3142–3155, 2017.