



**Please cite the Published Version**

Huang, Chenxi , Hong, Dan, Yang, Chenhui, Cai, Chunting, Tao, Siyi, Clawson, Kathy and Peng, Yonghong  (2023) A new unsupervised pseudo-siamese network with two filling strategies for image denoising and quality enhancement. *Neural Computing and Applications*, 35 (31). pp. 22855-22863. ISSN 0941-0643

**DOI:** <https://doi.org/10.1007/s00521-021-06699-9>

**Publisher:** Springer Science and Business Media LLC

**Version:** Published Version

**Downloaded from:** <https://e-space.mmu.ac.uk/634915/>

**Usage rights:**  [Creative Commons: Attribution 4.0](https://creativecommons.org/licenses/by/4.0/)

**Additional Information:** The version of record of this open access article, first published in *Neural Computing and Applications*, is available online at Publisher's website: <http://dx.doi.org/10.1007/s00521-021-06699-9>

**Data Access Statement:** Data of ADNI are a public database, and the URL is <http://adni.loni.usc.edu/>.

**Enquiries:**

If you have questions about this document, contact [openresearch@mmu.ac.uk](mailto:openresearch@mmu.ac.uk). Please include the URL of the record in e-space. If you believe that your, or a third party's rights have been compromised through this document please see our Take Down policy (available from <https://www.mmu.ac.uk/library/using-the-library/policies-and-guidelines>)



# A new unsupervised pseudo-siamese network with two filling strategies for image denoising and quality enhancement

Chenxi Huang<sup>1</sup> · Dan Hong<sup>1</sup> · Chenhui Yang<sup>1</sup> · Chunting Cai<sup>1</sup> · Siyi Tao<sup>1</sup> · Kathy Clawson<sup>2</sup> · Yonghong Peng<sup>3</sup>

Received: 21 April 2021 / Accepted: 27 October 2021 / Published online: 17 November 2021  
© The Author(s) 2021

## Abstract

Digital image noise may be introduced during acquisition, transmission, or processing and affects readability and image processing effectiveness. The accuracy of established image processing techniques, such as segmentation, recognition, and edge detection, is adversely impacted by noise. There exists an extensive body of work which focuses on circumventing such issues through digital image enhancement and noise reduction, but this work is limited by a number of constraints including the application of non-adaptive parameters, potential loss of edge detail information, and (with supervised approaches) a requirement for clean, labeled, training data. This paper, developed on the principle of Noise2Void, presents a new unsupervised learning approach incorporating a pseudo-siamese network. Our method enables image denoising without the need for clean images or paired noise images, instead requiring only noise images. Two independent branches of the network utilize different filling strategies, namely zero filling and adjacent pixel filling. Then, the network employs a loss function to improve the similarity of the results in the two branches. We also modify the Efficient Channel Attention module to extract more diverse features and improve performance on the basis of global average pooling. Experimental results show that compared with traditional methods, the pseudo-siamese network has a greater improvement on the ADNI dataset in terms of quantitative and qualitative evaluation. Our method therefore has practical utility in cases where clean images are difficult to obtain.

**Keywords** Pseudo-siamese network · Filling strategies · Image denoising · Attention

## 1 Introduction

Vision is an indispensable way for humans to obtain information and gain basic cognition and understanding of the world. However, the manual processing of large-scale

image datasets is both time-consuming and laborious. Efficiencies can be achieved through utilization of machine vision methods (for example, via automated monitoring, object/scene identification, and segmentation), but image quality is a key factor in machine vision performance. For example, in clinical medicine, medical imaging has become an important auxiliary tool for physicians. High-quality medical images can provide clear organ tissue and function information and improve the efficiency and accuracy of diagnosis and treatment. However, the image is inevitably injected with different concentrations and distribution of noise during the process of generation, storage, transmission, and application. The edges and characteristic information of the image are covered or easily blurred, resulting in the deterioration of the image quality, which does not meet the actual application requirements in production and scientific research.

---

✉ Chenxi Huang  
supermonkeyxi@xmu.edu.cn

✉ Chenhui Yang  
ych987@126.com

✉ Yonghong Peng  
y.peng@mmu.ac.uk

<sup>1</sup> School of Informatics, Xiamen University, Xiamen 361000, China

<sup>2</sup> School of Computer Science, St Peters Campus, Sunderland, UK

<sup>3</sup> Department of Computing and Mathematics, Manchester Metropolitan University, Manchester, UK

As the requirements for high-quality images continue to increase, denoising tasks have become a key branch in computer vision. The purpose of denoising is to remove image noise without losing critical information to the greatest extent possible and restore the latent clean image. Researchers have proposed different denoising methods, which are mainly divided into traditional [1–3] and deep learning methods.

Traditional methods include spatial filtering and frequency-based methods. Frequency domain methods transform the original signal into an easy pattern to denoise, such as a curvelet [4] and wavelet [5]. However, as far as wavelet transform is concerned, it fails to solve the problem of smooth edges. Spatial filtering is performed across a pixel's neighborhood region [1–3]. Methods include mean filtering [8], median filtering [9], Wiener filtering [10], and non-local mean filtering [11]. Mean filtering takes the average value around a pixel neighborhood instead of the original pixel. This method is simple but is a smoothing operation and therefore blurs edge and detail features. Median filtering, a similar neighborhood operation which utilizes the median value of neighboring pixels, preserves image definition but fails to suppress specific noise. Non-local mean filtering cleverly uses the redundant information of the image to retain image sharpness and detail, but has a higher computational complexity. Anisotropic filtering [12, 13] overcomes edge blurring associated with Gaussian filtering and maintains the image edge, but generally speaking, traditional methods are simple, require parameterization, are lacking in generalization, and are prone to blur or smoothing problems.

In light of the limitations of low-level traditional techniques, there has been recent focus on the application of deep learning for image enhancement and noise reduction. Deep learning has demonstrated superior performance in image denoising when compared with traditional techniques, across a variety of noise distributions. Methods include DnCNN [14], IRCNN [15], FFDNet [16], VST-net [17], and RED30 [18]. Supervised learning requires clean images as labels to guide training, and it can be very difficult to collect a sufficiently large number of clean labeled images in fields such as medicine and biology due to instrument or cost constraints [19]. As a result, a considerable part of supervised learning can only be used in synthetic images and may not generalize well when applied to real image contexts.

Unsupervised learning avoids this burden, instead extracting the structural characteristics in the noisy data itself. Lehtinen J proposes the Noise2Noise [20], which does not require clean images and directly uses independent noise image pairs. The denoising performance is close to the supervised learning methods. This method requires two images with the same content and independent noise.

In real scenes, it is not enough to obtain the noisy image pair [20], which greatly limits its practical application. To address these issues, Noise2Void [21], Noise2Self [22], Noise2Same [23], Noise2Sim [24], Self2Self [25], Deep Image Prior [26], 4D deep image prior [27] and convolutional blind spot neural networks [28] have been proposed. Noise2Void does not require noise image pairs and clean images as the target. It employs a blind spot network to predict the pixel and restores clean pixels from a single noise image by using neighboring pixel values. However, this method is not fully trained because of blind spots and makes prior assumptions about images and noise. If the conditions are not met, it will lead to poor denoising effects [28]. Inspired by non-local mean filtering, Noise2Sim utilizes the self-similarity of image to train the denoising network, which is innovative [29]. Noise2Atom [30] designs for scanning transmission electron microscopy images which takes advantage of two external networks, in order to apply additional constraints from the domain knowledge.

This paper builds upon the principle of Noise2Void, by proposing a pseudo-siamese network which utilizes noisy images directly without clean labels and noisy image pairs. The network fills blind spots of the image with two strategies. Specifically, one branch is directly filled with zeros, and the other uses random surrounding pixels. Different filling strategies construct different network inputs, and the two branches are, respectively, mapped to the new space to produce a representation. It is worth mentioning that the loss function designed in this paper consists of three parts, including the MSE loss of the prediction by the two branches and noise images, and the MSE loss between the two branches. Then, through the calculation of loss function, the similarity of the two inputs is narrowed to make it close to the potential clean image. Besides, the diverse filling strategy further avoids the problem of constant change. Although the parameters of the two branches are not shared, this paper improves the effective channel attention module [31] to communicate the two branches and adds the branch structure of the global maximum pooling.

## 2 Related work

### 2.1 Noise2Void

We use the formula  $x = s + n$  [32] to describe the generation of noise images.  $x$  represents the noise image, and  $s$  stands for the signal image without noise.  $n$  is the noise. From the noise generation process, the task of image denoising is deduced to separate the noise image  $x$  into two components: the clean signal  $s$  and noise  $n$ . We

subsequently represent a noisy image by the following joint distribution.

$$p(s, n) = p(s)p(n|s) \quad (1)$$

And assume  $p(s)$  to satisfy Formula 2 and it is an arbitrary distribution in Noise2Void. In other words, the pixels  $s_i$  do not satisfy statistically independent in signals [21]:

$$p(s_i|s_j) \neq p(s_i) \quad (2)$$

$s_i$  and  $s_j$  represent two pixels in images within a certain radius distance. In terms of the noise  $n$ , Noise2Void assumes that it satisfies the condition independent and zero-mean.

In order not to learn identity mapping, there is a special receptive field in the network that excludes the pixel of blind spot itself. In other words, the blind spot makes the predicted pixel value be affected by all the values of the rectangular neighborhood except the blind pixel  $x_i$  at position  $i$ . In training, the network can be expressed as following:

$$\arg \min_{\theta} \sum_j \sum_i L(f(x_{RF(i)}^j; \theta) = \hat{s}_i^j, s_i^j) \quad (3)$$

Here, the denoising network is regarded as a function  $f$ , where  $x_{RF(i)}^j$  is a patch around  $i$ , and the output is the pixel  $\hat{s}_i^j$ .  $\theta$  denotes the parameter of denoising network, and  $s_i^j$  is the target pixel corresponding to the input.  $L$  stands for the standard MSE loss.

In summary, Noise2Void opens the doors to a plethora of applications without large-scale clean images. Compared with the supervised algorithm, although it can achieve good denoising effects, it requires clean images to guide training, which greatly limits the practical application. Intuitively, N2V cannot be expected to perform better than supervised learning methods, but experiments show that it just drops in moderation. However, Noise2Void fails to make full use of the blind spot value for training and is inefficient [32]. There is still a certain gap between the results of Noise2Void denoising and the requirements of reality on the image.

## 3 Method

### 3.1 Pseudo-siamese network

The overall framework of the pseudo-siamese network proposed in this paper is shown in Fig. 1. The blind pixels in input images are randomly selected, in which one branch is filled with zeros and the other branch randomly selects neighboring pixels. Each branch utilizes the residual

network to predict the blind pixels and employs channel attention to connect the two branches and then weigh the importance of different channels. At the end of the network, a similarity calculation is used to make the results of the two branches closer to each other, with the aim of staying close to potentially clean images and preventing one branch from going in the wrong direction. During the training and testing of the pseudo-siamese network, clean images are not required; thus, it shows great potential in scenes such as medical and biological where it is difficult to obtain clean images.

According to the structure of pseudo-siamese network in Fig. 1, the input is noisy images, which is defined as  $N$ . The outputs of the two branches are denoted by  $D, D'$ , respectively. Each module is recorded as a function  $f$ ; hence, the processing of the pseudo-siamese network is represented as

$$D = f_{\text{conv1}}(f_{\text{att}}(f_{\text{res1}}(f_{\text{bsfn}}(N)))) \quad (4)$$

where  $f_{\text{conv1}}, f_{\text{att}}, f_{\text{res1}}, f_{\text{bsfn}}$  denote the functions of convolution, attention, resblocks, blind spot filling module in first branch with neighboring pixel, respectively. It is worth noting that for testing, this paper uses this branch.

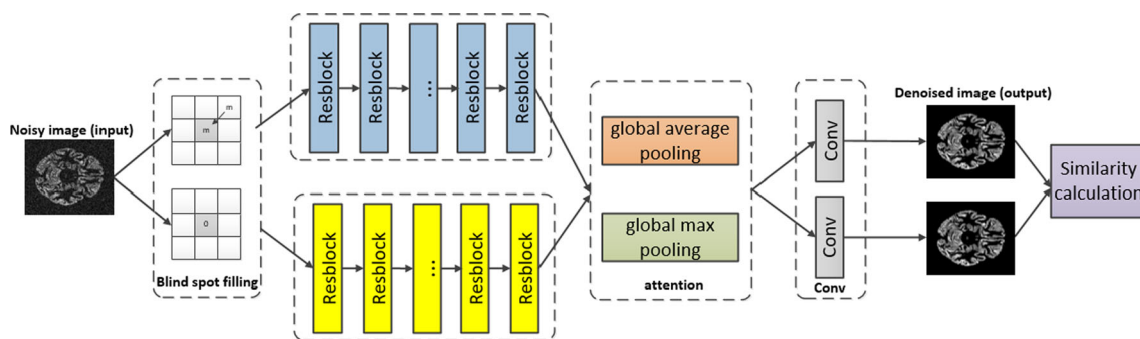
Similarly, a branch network filled with zero is represented as follows

$$D' = f_{\text{conv2}}(f_{\text{att}}(f_{\text{res2}}(f_{\text{bsfz}}(N)))) \quad (5)$$

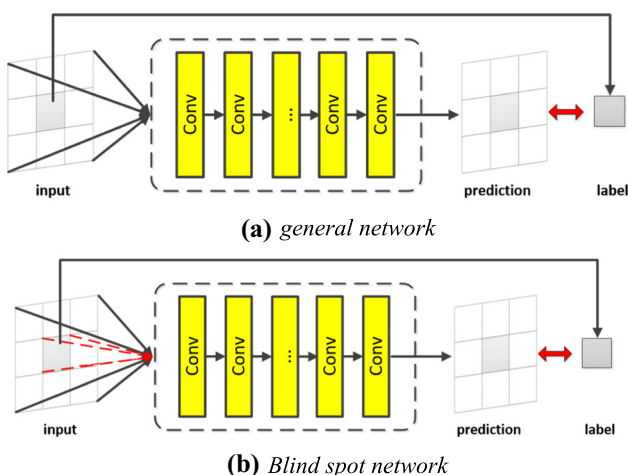
where  $f_{\text{conv2}}, f_{\text{att}}, f_{\text{res2}}, f_{\text{bsfz}}$  are defined the functions of convolution, attention, resblocks, blind spot filling module in second branch with zero, respectively. According to Formula 4 and 5, several observations can be made. Filling module and resblocks are not shared. In both branches, the parameters of the attention mechanism are shared. This paper also conducted experiments on shared parameters in resblocks, but the result is worse than that of separation. In addition, the attention mechanism is to communicate the two branches. On the whole, the pseudo-siamese network is both separate and connected.

### 3.2 Blind spot filling module

For a general network using exactly the same noisy image as input and label, the network will easily learn the identity mapping and cannot effectively extract image information, as shown in Fig. 2a. Therefore, based on Noise2Void, each branch of the pseudo-siamese network utilizes a blind spot to avoid identity mapping. Specifically, the network uses other values to replace the blind spot and then employs the information of rectangular neighborhood to predict, as shown in Fig. 2b. Although the input and target in the network are essentially noisy images, the network assumes that noise is condition independent and that neighboring information cannot predict the noise; hence, the hidden



**Fig. 1** The framework of the pseudo-siamese network, including convolution module, attention module, resblocks module and blind spot filling module. The parameters of resblocks in two branches are not shared but are composed of the same structure of residual modules



**Fig. 2** The general network and the blind spot network. The general network takes the entire noisy image as input, and the blind spot network replaces the blind spot pixels as input to avoid identity mapping

clean pixel values are extracted. As the network training converges, each branch eventually learns to remove the noise on the pixel and obtain clean images. The pixel values output by the two branches are constantly similar, so both effectively approach the direction of a clean image.

**3.3 Attention module**

Recently, numerous building modules have improved the achievable accuracy of deep learning and have shown extraordinary potential, such as attention [33], dilated convolution [15], memory block [34], and wavelet transform [35]. Such endeavors have been at the expense of computational efficiency. Based on the SENet [36] module, Efficient Channel Attention is lightweight with high efficiency and adaptively determines the size of the one-dimensional convolution kernel through the channel dimension. However, only global average pooling [37] is used in ECA. This paper adds the branch of global maximum pooling to extract more diverse feature textures, as

illustrated in Fig. 3. After sum the results of the global average pooling and global max pooling, the sigmoid is performed, and finally, element-wise product is used with the input.

Attention communicates two branches that do not share parameters and gives different weights to different channels to train them in a targeted way. On the one hand, the attention mechanism in this paper automatically adjusts the weight of each channel through training, so as to enhance the role of useful channels and suppress the irrelevant channels; on the other hand, it extracts more diverse channels through the global maximum pooling of another branch, which improves image denoising effect.

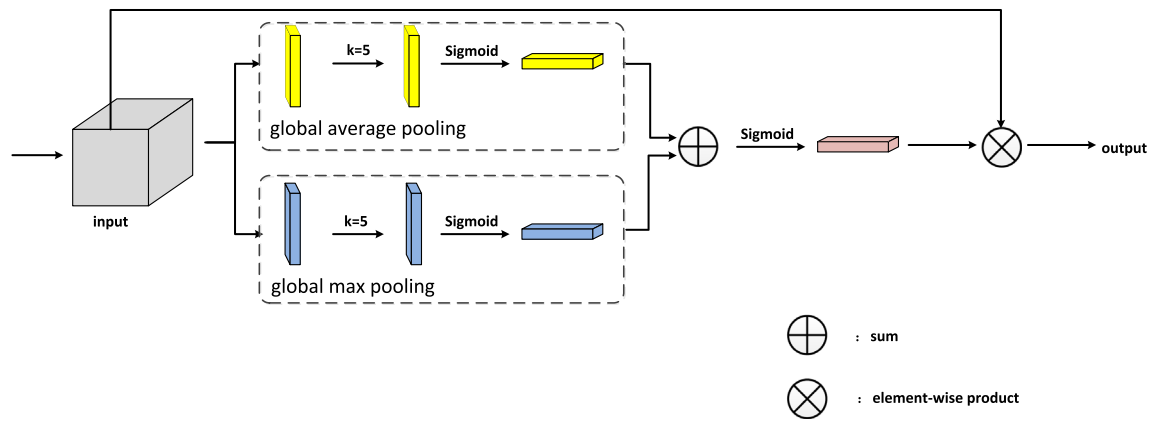
**3.4 Loss function**

The loss function designed in the paper is defined as follows

$$\arg \min_{\theta} \left( \begin{matrix} 5 * L(f(x_{neighbor}; \theta), y) + \\ 2 * L(f(x_{zero}; \theta), y) + \\ 3 * L(f(x_{zero}; \theta), f(x_{neighbor}; \theta)) \end{matrix} \right) \tag{6}$$

$x_{neighbor}$  represents the blind spot values replaced by random neighboring pixels.  $x_{zero}$  is the blind spot values directly replaced by zero.  $y$  stands for the label images, which is the original noise images.  $\theta$  denotes the parameters of pseudo-siamese network.

The loss function in this paper consists of three parts, including the loss between predictions of the two branches and label, and the similarity between the two branches. Experiments compare the three parts of the loss with the same weight and different weights. The results show that the PSNR of different weights is higher than the same. Therefore, we multiply the first branch and the second branch by 5 and 2, respectively, and compare the loss of similarity between two branches with weight 3. The weight is assigned mainly considers the first branch will be utilized to test, while the second branch is auxiliary. In addition, the



**Fig. 3** Add the branch of global maximum pooling to extract more diverse feature textures in attention module

similarity of the two branches is compared so that each branch can tend to clean labels during training.

Defining the pseudo-siamese network as a function  $f$ ,  $y_{neighbor}$  and  $y_{zero}$ , and representing the output of branch with  $x_{neighbor}$  and  $x_{zero}$ , respectively,  $\theta_{neighbor}$  and  $\theta_{zero}$  denote the parameters of two branches, then

$$\begin{aligned} f(x_{neighbor}; \theta_{neighbor}) &= y_{neighbor} \\ f(x_{zero}; \theta_{zero}) &= y_{zero} \end{aligned} \tag{7}$$

In the pseudo-siamese network, the mean square error (MSE) loss is utilized. The number of samples is  $m$ . The prediction result and label are denoted as  $y_{result}, y_{label}$ , respectively. The loss  $L$  is defined as follows:

$$L(y_{result}, y_{label}) = \frac{1}{m} \sum_{i=1}^m (y_{result}^{(i)} - y_{label}^{(i)})^2 \tag{8}$$

### 4 Experiment and analysis

This paper uses PSNR to objectively analyze the experimental results and displays the denoised images for intuitive visual inspection. For comparison, we investigate the performance of the pseudo-siamese network with BM3D [38], non-local mean filtering, median filtering, mean filtering, and Noise2Void across a variety of noise levels.

#### 4.1 Magnetic resonance imaging data

With the continuous development of artificial intelligence and image processing techniques, magnetic resonance imaging (MRI) has been widely adopted in medicine. The superiority of MRI [39–43] is primarily owned to non-radiation, non-invasiveness, and high resolution. The effective acquisition of original three-dimensional cross section imaging and multi-directional images without

reconstruction is also the preponderance of MRI. Compared with computed tomography (CT) [44], MRI has a significant advantage in the clarity of details of the central nervous system, joints, muscles, and other parts.

MRI the equipment captures images by collecting  $k$ -dimensional spatial data and performing Fourier transform [45]. The magnetic resonance coil of MRI equipment contains real part and imaginary part signals, and the phase difference between them is 90 degrees. Both real and imaginary signals contain additive white Gaussian noise with the mean value of zero, the variance of which is the same and independent.

According to the noise distribution, it has been demonstrated that MRI obeys Rician distribution [46–49]. The signal-to-noise ratio (SNR) [50], which refers to the ratio of the power spectrum of the signal-to-noise, is an important condition to estimate the noise distribution of MRI. In low SNR, Rician noise presents Rayleigh distribution [51], while it obeys Gaussian distribution in high SNR. Therefore, the complexity and variety of Rician noise increase the difficulty of denoising and make it become a huge challenge. To the best of our knowledge, few unsupervised learning methods have been applied to MRI denoising. Consequently, the research is of practical significance and theoretical value.

All experiments in the paper use the Alzheimer’s disease neuroimaging initiative dataset (ADNI), which is a public real brain dataset. We obtain a total of 199 MRI three-dimensional images, and each three-dimensional image is sliced along the axial plane. The images numbered 37–86 were selected to add Rician noise with noise level 10, 20, and 30, respectively. In all, there are a total of 9750 two-dimensional images, which are divided into 7750 training sets, 1000 test sets, and 1000 verification sets. All images are  $145 \times 121$  in size and are grayscale. During the experiment, Adam [52] is used for optimization, and the window setting size is  $5 \times 5$ .

## 4.2 Results analysis and discussion

### 4.2.1 Qualitative metrics

The experiment is carried out on Ubuntu 18.04, and the deep learning framework utilizes Pytorch 1.1.0. in Python. For experimental acceleration, NVIDIA GeForce GTX 1060 is used in the experiments.

In the field of image denoising, the commonly used quantitative evaluation is the peak signal-to-noise ratio (PSNR), which calculates the degree of distortion between a denoised image  $p$  and a clean image  $q$ . PSNR is the ratio between the maximum possible power of a signal and the power of corrupted noise that affects the fidelity of its representation [53, 54]. The larger the value, the smaller the distortion. The unit is dB. For a given  $M \times N$  image, PSNR is defined as follows:

$$\text{PSNR}(p, q) = 10 \log_{10} \frac{255^2 \times M \times N}{\sum_{(x,y) \in \Omega} |p(x,y) - q(x,y)|^2} \quad (9)$$

Experimental results of PSNR are shown in Table 1. Compared with the traditional methods based on transform domain and filtering, the pseudo-siamese network proposed in this paper greatly improves the denoising performance under various noise levels. For BM3D, our method has an improvement of more than 6 dB at all different noise levels; when the noise level is 30, the improvement can reach up to 7.75 dB. Compared with all the traditional methods, NLM achieves better denoising effects. In the noise level of 10 and 20, NLM has the highest PSNR, indicating that it is closest to a clean image in most traditional methods. However, compared with the pseudo-siamese method, NLM still has a large gap. When the noise level is 30, the gap of PSNR is 7.85 dB. Results clearly demonstrate the effectiveness of the pseudo-siamese network and are higher than, or comparable to, traditional methods.

It is worth mentioning that the traditional method has a poor denoising effect, while the two deep learning methods can reach more than 25 dB. In terms of the reference

**Table 1** PSNR of different methods. The best PSNR at different noise levels is highlighted in bold

Noise level $\lambda$	10	20	30
BM3D	24.88	20.54	17.79
NLM	26.76	21.05	17.69
Median filtering	24.62	20.84	18.33
Mean filtering	20.48	18.72	17.08
Noise2Void	30.72	27.54	25.44
Pseudo-siamese network(ours)	<b>31.16</b>	<b>28.04</b>	<b>25.54</b>

algorithm Noise2Void, when the noise level is 20, the performance of this paper is improved by 0.50 dB, but the two are equivalent under the 30 noise level, in which the difference is only 0.10 dB.

In order to illustrate the effectiveness of each module, ablation experiments are carried out based on the above settings. From Table 2, compared to Noise2Void at noise level 10, the pseudo-siamese network predicts clean images more effectively, which increases 0.14 dB. And attention with global maximum pooling branch increases 0.30 dB, which indicates that the improved attention extracts feature more fully.

### 4.2.2 Quantitative metrics

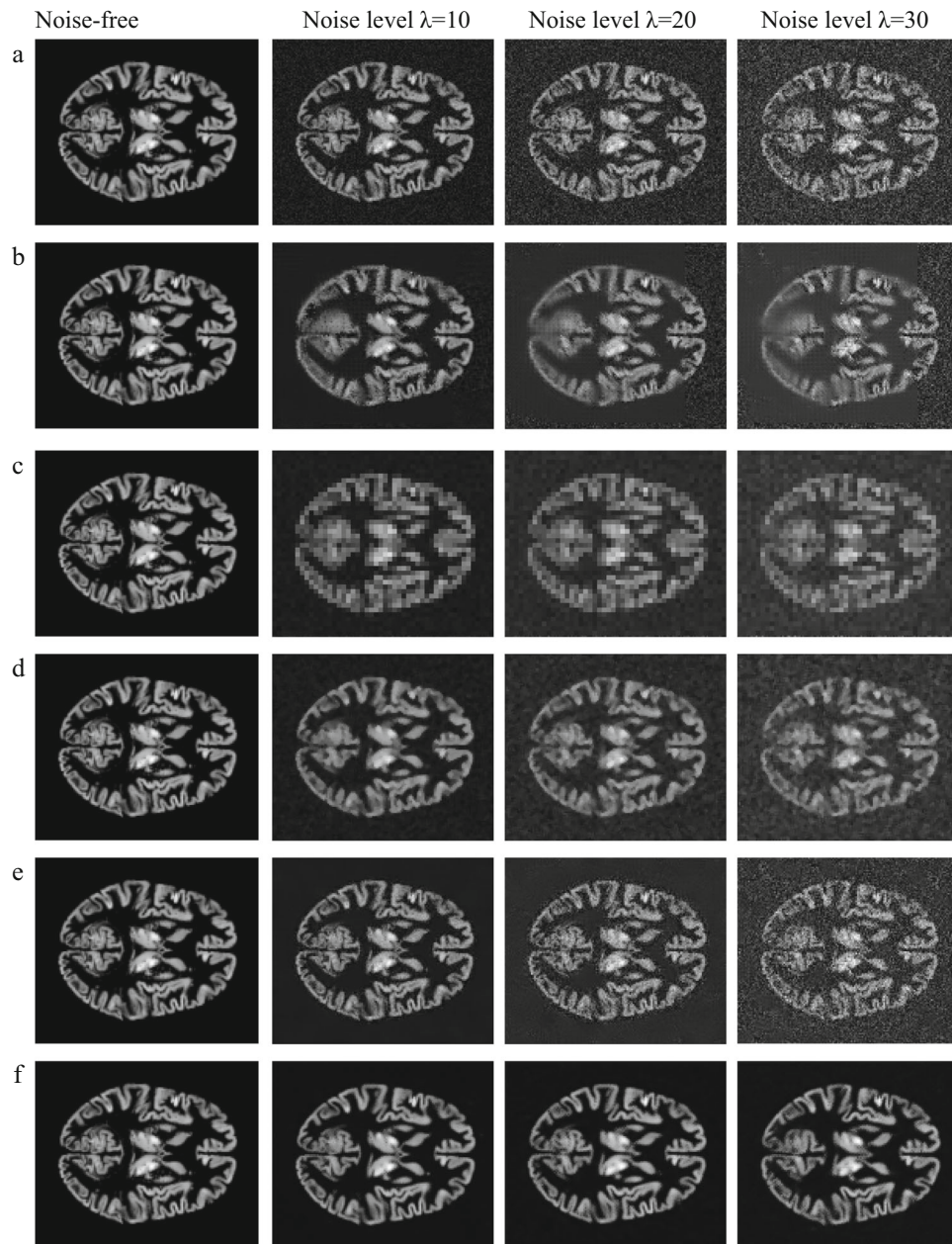
The PSNR as an objective evaluation criterion is with certain limitations. Therefore, in some cases, the value of PSNR is higher, but the visual effect on the image is poor. In view of above, this paper visually compares the denoised images, as shown in Fig. 4.

It can be seen from Fig. 4 that the traditional method has poor denoising effects on MRI with high noise levels. In contrast to clean labels, BM3D has a better recovery effect on image details than mean filtering and median filtering, but it causes a large-scale blur in the left half of MRI. In addition, it is similar to NLM when the level is 30, background noise cannot be effectively removed, and the denoising effect of NLM at noise level 30 cannot be seen intuitively. For the mean filtering and median filtering, they are with a certain denoising effect. However, the key details of the image are greatly blurred, which resulting in a decrease in image quality, and at a more extreme noise level of 30, the brain edges can no longer be clearly observed. As a whole, it is easy to blur the images and lose most of the tissue details with traditional methods.

In contrast, the pseudo-siamese network proposed in this paper can effectively eliminate noise and restore the original information better. Besides, the edge parts are relatively sharper, and the processing of background is also cleaner than traditional methods. In summary, compared with other experimental methods, the proposed pseudo-siamese network achieves the best effects in both qualitative and quantitative metrics.

**Table 2** PSNR of ablation experiments. The best PSNR at noise level 10 is highlighted in bold

Noise level $\lambda$	10
Noie2Void	30.72
Pseudo-siamese network	30.86
Pseudo-siamese network + ECA (ours)	<b>31.16</b>



**Fig. 4** Visual inspections denoised by traditional methods. **a** Images with different noise levels. **b** The results of BM3D. **c** The results of mean filtering. **d** The results of median filtering. **e** The results of NLM. The results of our pseudo-siamese network. **f** The results of our pseudo-siamese network

## 5 Conclusion

It is costly to obtain a large number of clean images and noisy pairs in a real image scene. Therefore, the application of supervised methods in this domain is limited. Inspired by the blind spot network Noise2Void, this paper designs a new pseudo-siamese network and combines channel attention with only noisy images. As part of this, the network employs two different strategies to fill the blind spot

for different branches which fill neighboring pixel values and zeros, respectively. On the one hand, the neighboring pixel features are utilized to predict corresponding blind spots, and on the other hand, the results of the two branches are similar to the final clean image. This paper conducts experiments to verify the effectiveness of the pseudo-siamese network. Compared with traditional methods under three noise levels, the performance of ours has been greatly improved in terms of qualitative metrics and quantitative



metrics. In summary, the method reduces the distortion of the output denoised image with noisy images, which opens a door to medical, biological, and other fields.

However, the modules added have increased the complexity and calculations in the network to a certain extent. In the later research, the trade-off between efficiency and effectiveness will be more considered, and if the noise distribution does not meet the preset of Noise2Void, how to improve the effect in denoising will be studied further.

**Acknowledgements** This work was supported by the National Natural Science Foundation of China (Grant No. 62002304) and the Natural Science Foundation of Fujian Province of China (Grant No. 2020J05002). The authors would like to thank the Alzheimer's Disease Neuroimaging Initiative (ADNI) for providing data for this paper.

**Data Availability** Data of ADNI are a public database, and the URL is <http://adni.loni.usc.edu/>.

## Declarations

**Conflict of interest** The authors declare that they have no conflict of interest.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Yaroslavsky LP (2012) Digital picture processing: an introduction[M]. Springer Science & Business Media
- Smith SM, Brady JM (1997) SUSAN—a new approach to low level image processing[J]. *Int J Comput Vision* 23(1):45–78
- Tomasi C, Manduchi R (1998) Bilateral filtering for gray and color images[C]. In: Sixth international conference on computer vision (IEEE Cat. No. 98CH36271). IEEE, 839–846
- Ma J, Plonka G (2010) The curvelet transform[J]. *IEEE Signal Process Mag* 27(2):118–133
- Weaver JB, Xu Y, Healy DM Jr et al (1991) Filtering noise from images with wavelet transforms[J]. *Magn Reson Med* 21(2):288–295
- Wang X, Zhang W, Li R et al (2019) The UDWT image denoising method based on the PDE model of a convexity-preserving diffusion function[J]. *EURASIP J Image Video Process* 2019(1):1–9
- Anand CS, Sahambi JS (2010) Wavelet domain non-linear filtering for MRI denoising[J]. *Magn Reson Imaging* 28(6):842–861
- Changlai G (2007) Image-denoising method based on wavelet transform and mean filtering[J]. *Opto-Electron Eng* 1:19
- Boyat A, Joshi BK Image denoising using wavelet transform and median filtering[C]. In: 2013 Nirma University international conference on engineering (NUI-CONE). IEEE, 2013: 1–6
- Kazubek M (2003) Wavelet domain image denoising by thresholding and Wiener filtering[J]. *IEEE Signal Process Lett* 10(11):324–326
- Buades A, Coll B, Morel JM (2011) Non-local means denoising[J]. *Image Process On Line* 1:208–212
- Perona P, Malik J (1990) Scale-space and edge detection using anisotropic diffusion[J]. *IEEE Trans Pattern Anal Mach Intell* 12(7):629–639
- Catté F, Lions PL, Morel JM et al (1992) Image selective smoothing and edge detection by nonlinear diffusion[J]. *SIAM J Numer Anal* 29(1):182–193
- Zhang K, Zuo W, Chen Y et al (2017) Beyond a gaussian denoiser: residual learning of deep cnn for image denoising[J]. *IEEE Trans Image Process* 26(7):3142–3155
- Zhang K, Zuo W, Gu S, et al. (2017) Learning deep CNN denoiser prior for image restoration[C]. In: Proceedings of the IEEE conference on computer vision and pattern recognition, 3929–3938
- Zhang K, Zuo W, Zhang L (2018) FFDNet: Toward a fast and flexible solution for CNN-based image denoising[J]. *IEEE Trans Image Process* 27(9):4608–4622
- Zhang M, Zhang F, Liu Q et al (2019) VST-net: variance-stabilizing transformation inspired network for Poisson denoising[J]. *J Vis Commun Image Represent* 62:12–22
- Mao XJ, Shen C, Yang Y (2016) Image restoration using very deep convolutional encoder-decoder networks with symmetric skip connections[J]. *arXiv preprint arXiv:1603.09056*.
- Huang C, Zong Y, Chen J et al (2021) A deep segmentation network of stent struts based on IoT for interventional cardiovascular diagnosis[J]. *IEEE Wirel Commun* 28(3):36–43
- Lehtinen J, Munkberg J, Hasselgren J, et al. (2018) Noise2noise: learning image restoration without clean data[J]. *arXiv preprint arXiv:1803.04189*
- Krull A, Buchholz T O, Jug F (2019) Noise2void-learning denoising from single noisy images[C]. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2129–2137
- Batson J, Royer L (2019) Noise2self: Blind denoising by self-supervision[C]. In: International conference on machine learning. PMLR, 524–533
- Xie Y, Wang Z, Ji S (2020) Noise2Same: optimizing a self-supervised bound for image denoising[J]. *arXiv preprint arXiv:2010.11971*
- Niu C, Wang G. (2020) Noise2Sim—similarity-based self-learning for image denoising[J]. *arXiv preprint arXiv:2011.03384*
- Quan Y, Chen M, Pang T, et al. (2020) Self2self with dropout: learning self-supervised denoising from single image[C]. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 1890–1898
- Ulyanov D, Vedaldi A, Lempitsky V. (2018) Deep image prior[C]. In: Proceedings of the IEEE conference on computer vision and pattern recognition, 9446–9454
- Hashimoto F, Ohba H, Ote K et al (2021) 4D deep image prior: Dynamic PET image denoising using an unsupervised four-dimensional branch convolutional neural network[J]. *Phys Med Biol* 66(1):015006
- Laine S, Karras T, Lehtinen J, et al. High-quality self-supervised deep image denoising[J]. *arXiv preprint arXiv:1901.10277*, 2019
- Song J, Jeong JH, Park DS et al (2020) Unsupervised denoising for satellite imagery using wavelet directional CycleGAN[J]. *IEEE Trans Geosci Remote Sens* 59:6823–6839

30. Wang F, Henninen TR, Keller D et al (2020) Noise2Atom: unsupervised denoising for scanning transmission electron microscopy images[J]. *Appl Microsc* 50(1):1–9
31. Qilong W, Banggu W, Pengfei Z, et al. (2020) ECA-Net: efficient channel attention for deep convolutional neural networks[J]
32. Wu X, Liu M, Cao Y, et al. Unpaired learning of deep image denoising[C]. In: *European Conference on computer vision*. Springer, Cham, 2020: 352–368
33. Anwar S, Barnes N (2019) Real image denoising with feature attention[C]. In: *Proceedings of the IEEE/CVF international conference on computer vision*. 3155–3164
34. Tai Y, Yang J, Liu X (2017) Image super-resolution via deep recursive residual network[C]. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, 3147–3155
35. Liu P, Zhang H, Zhang K, et al. (2018) Multi-level wavelet-CNN for image restoration[C]. In: *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, 773–782
36. Hu J, Shen L, Sun G (2018) Squeeze-and-excitation networks[C]. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, 7132–7141
37. Roig C, Varas D, Masuda I, et al (2021) Generalized local attention pooling for deep metric learning[C]. In: *2020 25th International conference on pattern recognition (ICPR)*. IEEE, 9951–9958
38. Dabov K, Foi A, Katkovnik V et al (2007) Image denoising by sparse 3-D transform-domain collaborative filtering[J]. *IEEE Trans Image Process* 16(8):2080–2095
39. Jadvar H, Colletti PM (2014) Competitive advantage of PET/MRI[J]. *Eur J Radiol* 83(1):84–94
40. Jiang Y, Zhao K, Xia K et al (2019) A novel distributed multitask fuzzy clustering algorithm for automatic MR brain image segmentation[J]. *J Med Syst* 43(5):1–9
41. Ikram S, Shah JA, Zubair S et al (2019) Improved reconstruction of MR scanned images by using a dictionary learning scheme[J]. *Sensors* 19(8):1918
42. Tripathi PC, Bag S (2020) CNN-DMRI: a convolutional neural network for denoising of magnetic resonance images[J]. *Pattern Recogn Lett* 135:57–63
43. Yu H, Ding M, Zhang X (2019) Laplacian eigenmaps network-based nonlocal means method for MR image denoising[J]. *Sensors* 19(13):2918
44. Kapoor V, McCook BM, Torok FS (2004) An introduction to PET-CT imaging[J]. *Radiographics* 24(2):523–543
45. Bracewell RN, Bracewell RN (1986) *The Fourier transform and its applications*[M]. McGraw-Hill, New York
46. Nowak RD (1999) Wavelet-based Rician noise removal for magnetic resonance imaging[J]. *IEEE Trans Image Process* 8(10):1408–1419
47. Bhadauria HS, Dewal ML (2013) Medical image denoising using adaptive fusion of curvelet transform and total variation[J]. *Comput Electr Eng* 39(5):1451–1460
48. Li S, Zhou J, Liang D et al (2020) MRI denoising using progressively distribution-based neural network[J]. *Magn Reson Imaging* 71:55–68
49. He L, Greenshields IR (2008) A nonlocal maximum likelihood estimation method for Rician noise reduction in MR images[J]. *IEEE Trans Med Imaging* 28(2):165–172
50. Hoult DI, Richards RE (1976) The signal-to-noise ratio of the nuclear magnetic resonance experiment[J]. *J Magn Reson* 24(1):71–85
51. Kuruoglu EE, Zerubia J (2004) Modeling SAR images with a generalization of the Rayleigh distribution[J]. *IEEE Trans Image Process* 13(4):527–533
52. Kingma D P, Ba J. Adam: A method for stochastic optimization[J]. *arXiv preprint arXiv:1412.6980*, 2014.
53. Yousuf MA, Nobi MN (2011) A new method to remove noise in magnetic resonance and ultrasound images[J]. *J Sci Res* 3(1):81–81
54. Setiadi DRIM (2021) PSNR vs SSIM: imperceptibility quality assessment for image steganography[J]. *Multimed Tools Appl* 80:8423–8444

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.