

Article

Self-Supervised Joint Learning for pCLE Image Denoising [†]

Kun Yang , Haojie Zhang ^{*}, Yufei Qiu, Tong Zhai and Zhiguo Zhang

State Key Lab of Information Photonics and Optical Communications, Beijing University of Posts and Telecommunications (BUPT), Beijing 100876, China; ykk@bupt.edu.cn (K.Y.); qiuyufei@bupt.edu.cn (Y.Q.); zhai_tong@bupt.edu.cn (T.Z.); zhangzhiguo@bupt.edu.cn (Z.Z.)

^{*} Correspondence: haojie.zhang@bupt.edu.cn

[†] This paper is an extended version of our paper published in: A Self-Supervised Denoising Method for pCLE Images. In Proceedings of the 2023 8th International Conference on Image, Vision and Computing (ICIVC), Dalian, China, 27–29 July 2023.

Abstract: Probe-based confocal laser endoscopy (pCLE) has emerged as a powerful tool for disease diagnosis, yet it faces challenges such as the formation of hexagonal patterns in images due to the inherent characteristics of fiber bundles. Recent advancements in deep learning offer promise in image denoising, but the acquisition of clean-noisy image pairs for training networks across all potential scenarios can be prohibitively costly. Few studies have explored training denoising networks on such pairs. Here, we propose an innovative self-supervised denoising method. Our approach integrates noise prediction networks, image quality assessment networks, and denoising networks in a collaborative, jointly trained manner. Compared to prior self-supervised denoising methods, our approach yields superior results on pCLE images and fluorescence microscopy images. In summary, our novel self-supervised denoising technique enhances image quality in pCLE diagnosis by leveraging the synergy of noise prediction, image quality assessment, and denoising networks, surpassing previous methods on both pCLE and fluorescence microscopy images.

Keywords: probe confocal laser endomicroscopy; confocal; image denoising; self-supervised



Citation: Yang, K.; Zhang, H.; Qiu, Y.; Zhai, T.; Zhang, Z. Self-Supervised Joint Learning for pCLE Image Denoising. *Sensors* **2024**, *24*, 2853. <https://doi.org/10.3390/s24092853>

Academic Editor: Stefania Perri

Received: 2 April 2024

Revised: 26 April 2024

Accepted: 28 April 2024

Published: 30 April 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Probe-based confocal laser endomicroscopy (pCLE) is a technology that applies microscopic imaging techniques to endoscopic examinations in the medical field [1]. Optical scanning confocal microscopy is a mature technology that offers a significant advantage over traditional microscopic imaging, allowing for selective reduction of light from out-of-focus planes, thereby enabling depth imaging [2]. Clinical trials have demonstrated the value of this new technology. When utilizing pCLE, a laser beam is employed to illuminate the area of interest within biological tissues [3]. Currently, pCLE has found application across a wide range of biomedical scenarios, including the cavity structures of the digestive system, respiratory system, and urogenital system tissues or organs [4,5].

Due to the utilization of fiber bundles (FB) in pCLE, the spatial resolution of the FB imaging system is constrained by the core diameter and density of the optical fibers [6]. Additionally, variations in the light transmission properties of individual fibers and their surrounding sheaths result in the formation of honeycomb patterns, which superimpose on the imaging results and hinder the precise analysis of objects [7]. Over the past few years, various methods have been developed to eliminate honeycomb patterns, such as employing bandpass filters in the Fourier domain [8–11]. However, selecting an appropriate threshold in the frequency domain to eliminate honeycomb patterns without blurring the underlying imaging structures poses a significant challenge, which would severely affect image clarity [12–14]. Therefore, there is an urgent need to remove noise and enhance resolution.

In traditional deep learning, denoising methods typically involve neural networks learning the mapping relationship from noisy images to clean images [15]. This is a supervised learning approach, and to apply supervised learning within deep neural networks,

a sufficient amount of labeled data is required [16–19]. However, this method faces some bottlenecks in practical applications, particularly in the medical field, where obtaining true clean images for training data can be extremely challenging. The noise value of a noisy pixel not only exhibits spatial correlation with the surrounding noise but also demonstrates some correlation between different color channels within a single pixel [20]. Existing techniques mostly focus on removing Gaussian noise, while most medical images are functional images where factors such as a single light source, detection method, and human body thickness often lead to non-uniform noise distribution [21]. For example, the noise in pCLE images may vary with factors like different cameras and gastrointestinal environments making traditional deep learning denoising methods potentially limited in dealing with situations where obtaining true clean data is challenging as in medical images.

To address this challenge, many methods have incorporated noise-based prior information to handle degraded images [22–26]. One advantage of this approach is that it can handle specific levels of noise variance, as the network has encountered this type of noise during training [27]. However, they often struggle to generalize well to unseen noise levels or types. In the development of self-supervised denoising methods, several strategies have been proposed to train them exclusively on noisy images [28]. Krull introduced the first self-supervised method called Noise2Void [29], which utilizes the concept of blind spots to avoid the network learning a constant mapping from noisy images. The idea is to avoid this by randomly sampling pairs of masked and unmasked regions within the same image, represented as (Y_i^{masked}, Y_i) . However, the mask of blind spot network leads to the loss of high-frequency information, and this method assumes noise independence, making it less effective for structured noise. Batson [30] demonstrated that different masks can also influence denoising performance, and they replaced the mask with the average of surrounding pixels. Another approach, as presented in [31], involves designing masks based on masked pixels and adding a noise prediction network trained jointly with the denoising network to improve denoising performance. Recent research DIP [32] observed that the structure of the generator network is sufficient to capture image statistics before any learning, allowing it to be trained to recover images from reconstructions without requiring extensive datasets. This was the first study to directly investigate the priors captured by deep convolutional generative networks, rather than learning network parameters from images. Building on this research, ref. [33] demonstrated that a single noisy image itself can be used to train denoising networks with competitive performance. The results of single-image self-supervised learning not only provide a practical neural network-based image denoiser but also inspire further research into self-supervised learning in other image restoration problems [34,35]. Moreover, noise prediction plays an important role in self-supervised denoising, and the CDB-Net method of Feng et al. has shown good results in the field of noise prediction [36]. However, inherent limitations persist in the efficacy of information extraction from such predictions [37].

While utilizing noise prior information can offer effective solutions, there is a need for further research to enhance the method's generalization performance to accommodate unknown noise levels and types. To address this issue, we introduce an evaluation criterion throughout the entire training framework to utilize metrics for determining which image exhibits the best restoration quality. We propose a no-reference image quality assessment metric, and subsequently we trained a neural network to select the optimal denoised image during training. This assessment metric, combined with state-of-the-art denoising techniques, has been applied to denoise pCLE images and other fluorescence microscopy images lacking clear reference images. Our approach achieved significant improvements in denoising tasks for pCLE images and fluorescence microscopy datasets.

2. Materials and Methods

2.1. Architecture

The self-supervised approach can be considered as a specialized form of unsupervised learning that mimics supervised learning through self-imposed tasks, rather than relying on

predetermined prior knowledge. In contrast to fully unsupervised settings, self-supervised learning constructs pseudo-labels using information inherent in the dataset. The automatic acquisition of these pseudo-labels is crucial in self-supervised learning. In this study, we propose a self-supervised-constrained pCLE noise image denoising network.

This network consists of three scale branches, as illustrated in Figure 1. The overall architecture includes a non-blind denoising sub-network D-Net, a quality assessment sub-network Q-Net, and a noise prediction sub-network N-Net. In our previous work, we introduced a self-supervised learning method that incorporates an image quality assessment network to identify the optimal moments during training.

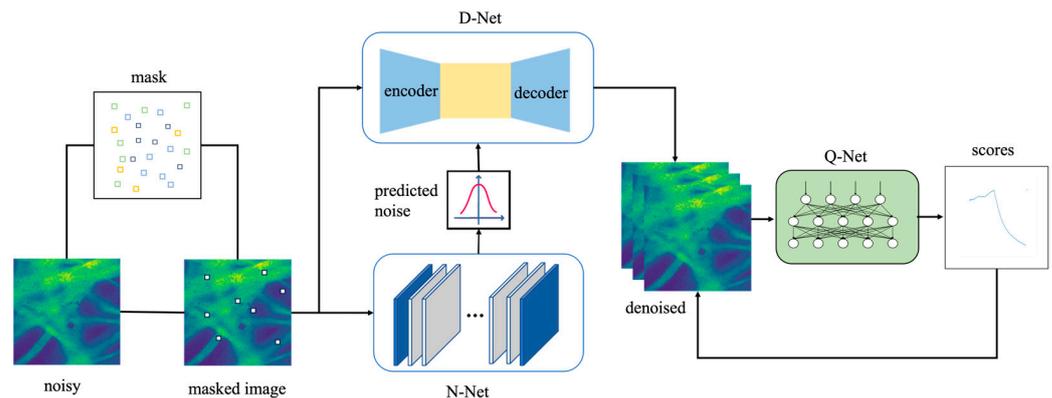


Figure 1. Architecture of proposed method.

The deep learning method for solving the image denoising problem can be transformed into training the function f_θ according to the unknown parameter θ . In order to prevent the image from learning a constant mapping to a noisy image, the image is processed before the denoising network with an added mask, and the training samples and the learning objects are the images after the added mask, respectively; they can be expressed as $\min_{\theta} \sum_m L(f_\theta(\hat{y}, y - \hat{y}))$. Since the self-supervised denoising algorithm we use has no clean data and uses only noisy images and masks, the goal is to minimize the self-supervised loss of the form $\theta \rightarrow \sum_{i=1}^N \|f_\theta(Y_i^{masked} - Y_i)\|_2^2$, where Y_{masked} is an image in which the pixel Y_i has been masked using M . Thus, for the image denoising network, the only a priori information about the network is currently the structure of the network itself, so N-Net is used to predict the image noise information as new image features to be trained along with the noisy image.

N-Net extracts noisy observations y to generate an estimated noise level map ($(\sigma(y) = F_E(y; W_E))$), where W_E represents the network parameters of N-Net. The output of N-Net is the noise level map, as it has the same dimensions as the input y and can be estimated through a fully convolutional network. Then, D-Net takes both y and $\sigma(y)$ as input and produces the final denoised result ($(x = F_D(y, \sigma(y); W_D))$). The noise network and the noise prediction network share the same parameters. The role of D-Net is to perform denoising under the guidance of noise intensity. N-Net's extracted noise information and noisy image are input into the denoising network D-Net. During the training process, as the number of iterations increases, the image tends to be reconstructed.

We use an advanced Quality Assessment Network called Q-Net to evaluate reconstructed images. This network assesses the quality of reconstructed images, generating a score trend map for each image. By using the results from Q-Net, we figure out the best restoration moment for each image. It is important to note that detailed explanations of both D-Net and Q-Net can be found in previous works. Therefore, this paper mainly focuses on providing a comprehensive and detailed explanation of the N-Net network.

2.2. Architecture of the Noise Prediction Network

To overcome the limitations of existing techniques, this invention leverages Convolutional Neural Networks with the aim of providing a more precise prediction of the variance of real noise and using this estimated noise variance to assist in image denoising. The technical approach employed in this invention is based on a deep learning method that combines noise prediction and image denoising. For a given noisy image, the image model is defined as $y = x + \epsilon$. Here, x represents the noisy image, y represents the original image, and ϵ represents the noise introduced during the camera's processing. This approach integrates noise prediction and image denoising into a joint framework, utilizing deep learning methods to enhance the quality of denoised images by accurately estimating the characteristics of the noise present in the captured images. N-Net is a fully convolutional network consisting of 20 convolutional layers and does not include pooling layers or batch normalization operations. Each convolutional layer, except for the output layer, is accompanied by a non-linear activation function. Its architecture is depicted in Figure 2.

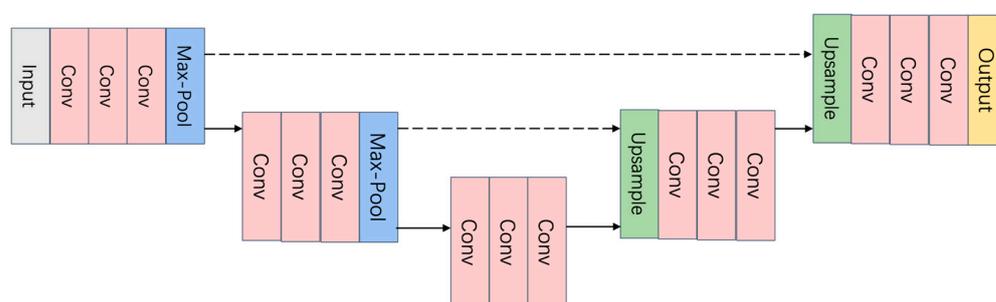


Figure 2. Architecture of the N-Net.

D-Net utilizes a U-shaped network framework, where the encoder employs max-pooling operations for downsampling, aiding in the exploration of multi-scale information within the image and expanding the feature receptive field. The decoder uses bilinear interpolation for upsampling, facilitating the reconstruction of denoised images. This setup ensures that both N-Net and D-Net are well-structured to perform their respective tasks in the image denoising process.

3. Results

3.1. Dataset

In the case of pCLE images where actual data is not available, our dataset is solely used for evaluating the visual quality of denoising. Thus, we train and evaluate our method on the Widefield2SIM (W2S) dataset, which is an openly accessible microscopy image dataset. W2S is one of the representative real-world datasets, containing well-aligned noisy-clean image pairs for training. We selected this dataset because we aimed to facilitate a fair comparison, utilizing the same dataset as used by the authors in [31,38].

In our experiments, W2S-1, W2S-2, and W2S-3 represent three channels of the W2S dataset [39], each with 120 fields of view (FOV), which can be treated as individual datasets. For each FOV, there are 400 noisy images with size of 512×512 pixels, and only one observation is used for training and evaluation. Across all three datasets, we generate the ground truth images using image averaging. The acquisition of noise-free images involves capturing 400 consecutive shots of a static scene and then taking the average. The noise labels are obtained by taking the difference between the noisy images and the noise-free images. Noise variance maps reflect the fluctuations across the 400 images with respect to the noise-free image. The standard deviation is computed from the variance map, which is then normalized.

3.2. Implementation

3.2.1. D-Net

In our experiments, we use the network parameters: A 3-depth U-Net with 1 input channel and 64 channels in the first layer; we use padding to the 3×3 convolution layer, meaning that we do not change the height and width of the feature layers [40]. Since the feature layer size is not changed after convolution, the two feature maps can be directly stitched together, so there is no need for center cropping, and the final obtained convolution layer height and width are kept the same as the input convolution layer height and width. The number of iterations is set to 3000, and the image is saved every 100 iterations.

3.2.2. Q-Net

Our implementation is based on the hyperIQA implementation [41–43]. Q-Net includes three parts of functions. The first part is the semantic extraction part of the framework based on ResNet50. The second part is the perceptual rule-building part; the input is the output of the last layer of ResNet of feature extraction. The third part of the quality prediction part of the network structure is relatively simple, with only four FC layers, but the parameter weights of these four FC layers are provided by the second part; the final output is the final score of the image.

3.2.3. N-Net

To set network parameters, we specify that the convolutional kernels are all of size 3×3 , and each convolutional layer has 64 kernels. To ensure that the image size remains unchanged after each convolution and to prevent boundary effects, we apply zero-padding during each convolution. The initial learning rate is set to 1×10^{-4} (0.0001). We employ the Adam optimization algorithm.

3.3. Evaluation on the Image Denoising Network

In order to test the performance of the denoising network and determine if it is consistent with the denoising process, where the images are first recovered and then noise is applied, we use the W2S dataset for image recovery and save the images within the number of iterations in order to calculate the PSNR and verify whether it is consistent with the trend of all getting better before getting worse in the recovery process.

Figure 3 shows the method of recovering a single image and the change of the image during the recovery process. It can be seen that the image gradually becomes clearer during the reconstruction process, but the trend of image quality change is not linear, and according to Figure 3, we can initially speculate that the image quality becomes better and then worse within a certain number of iterations.

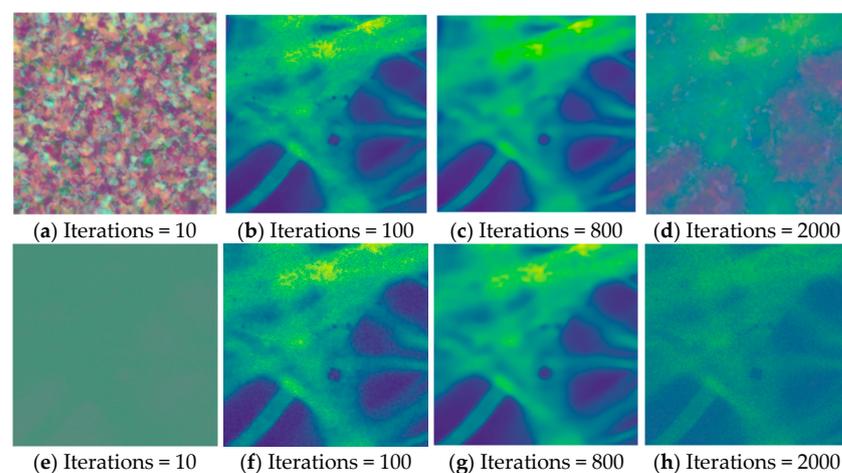


Figure 3. The reconstruction process of single image denoising. We show the reconstruction process of recovering a single image for DIP and S2S. Upper: DIP, Lower: Self2Self.

In order to further prove this point, the change trend graph is obtained as in Figure 4 by calculating the PSNR during the recovery process of the two datasets in the W2S dataset. It can be seen that there is still no upward trend in the two images after the number of iterations has been extended to 2800 times, and before that, the PSNR becomes higher and then lower, and the number of iterations is not the same for different datasets to reach the peak, based on which it can be concluded that the necessity of adding Q-Net.

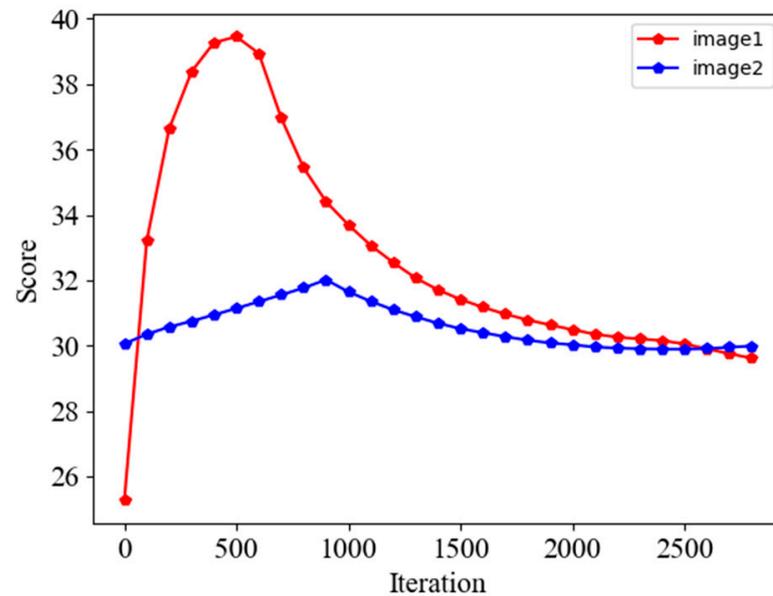


Figure 4. PSNR trends of the two datasets of W2S after passing through the denoising network.

3.4. Evaluation on the Image Quality Assessment Network

To evaluate the ability of Q-Net to assess image quality, we conducted an ablation study to quantify the performance contribution of Q-Net. Before that, this network was first evaluated and feasibility analysis was performed to generate a new dataset by adding different types and different intensities of synthetic noise to the acquired pCLE dataset images by noise addition process. We used three classical noise models: Gaussian noise, pretzel noise, and Poisson noise. The images with different noise intensities were scored separately, and the prediction scores of the datasets with different noises are shown in Figures 5–7.

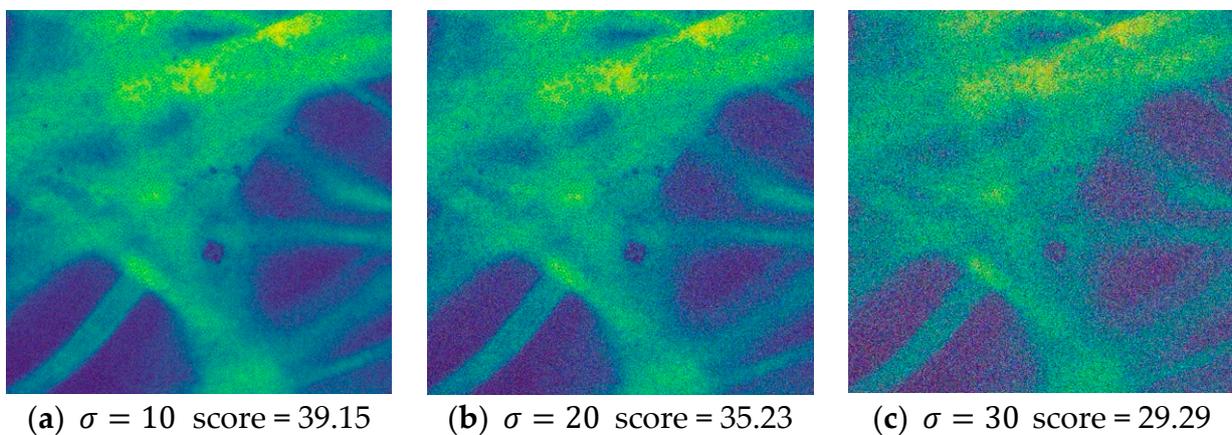


Figure 5. Quality assessment of the Gaussian noise images with different intensities.

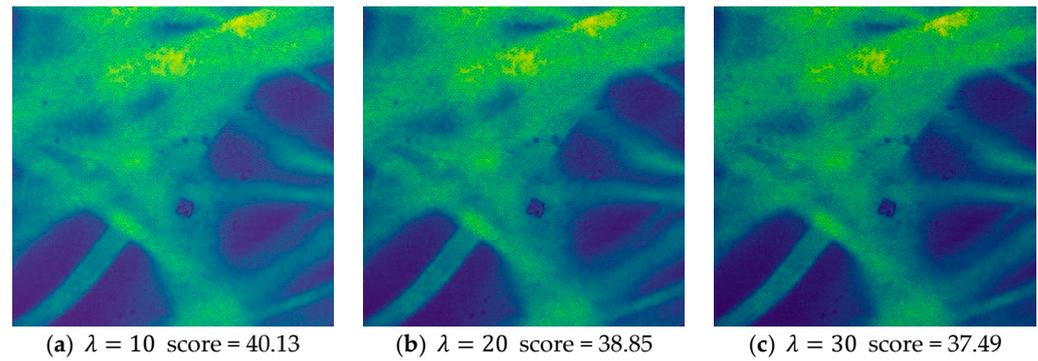


Figure 6. Quality assessment of the Poisson noise images with different intensities.

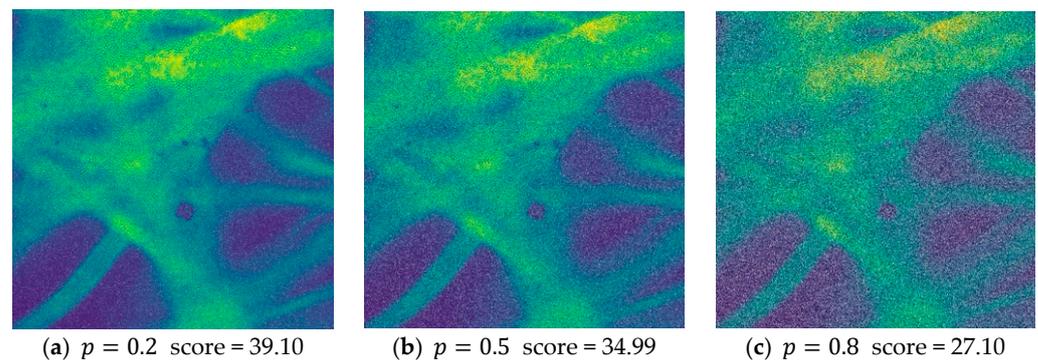


Figure 7. Quality assessment of the Pepper noise images with different intensities.

From this result, we can see that the predicted scores and the corresponding noisy images can be matched, and according to our visual subjective evaluation, it is obvious that the images with high noise intensity contain little information and possess lower scores. All three types of noise match this property, proving that the sub-IQA has good performance.

When applying IQA to the overall image denoising, we need the scores in the whole set of images and select the images with the highest scores to achieve the goal of improving image quality. We used two different pCLE images for the test, and the figure shows the score results for each image. It can be seen that the trend of the scores increases regularly and then decreases so that the image will be recovered best in a certain recovery process, as shown in Figure 8, and the figure also verifies this property. There is a certain difference in the optimal number of iterations for different images, which reflects that our need to add a quality assessment network is necessary; for example, if the number of iterations is set to 1200, image1 can achieve better results, but the other is not the best for recovery.

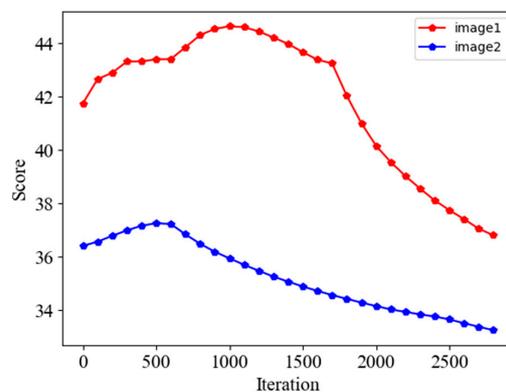


Figure 8. The trend of the scores of the two datasets of pCLE after passing through the two networks is increasing and then decreasing for all two images, and the number of iterations to reach the maximum score is different.

The above is the analysis of the Q-Net. For the next, we will conduct denoising experiments on the whole part to observe the effect.

3.5. Compare to Other Deep Learning Denoising Methods

We compared our method to three baselines: N2V, JOINT, which is the self-supervised blind denoising method that has shown the best results on the datasets we considered, and S2S, whose standard deviation is chosen to maximize the PSNR on the evaluation dataset. Figure 9 shows the effect of pCLE image denoising; we only compare the subjective visual quality due to the fact that pCLE does not exist in the real image. When evaluating the subjective quality and information content of an image, several factors come into play. Firstly, clarity and detail are paramount, with high-quality images exhibiting clear details and sharp edges. Additionally, color accuracy and contrast play crucial roles, as vibrant colors and appropriate contrast levels contribute to overall image quality. Furthermore, minimizing noise levels, particularly in darker areas and flat regions, is essential to maintain image integrity. Lastly, an image's richness in conveying intended content or message is vital, requiring sufficient detail to effectively communicate its information.

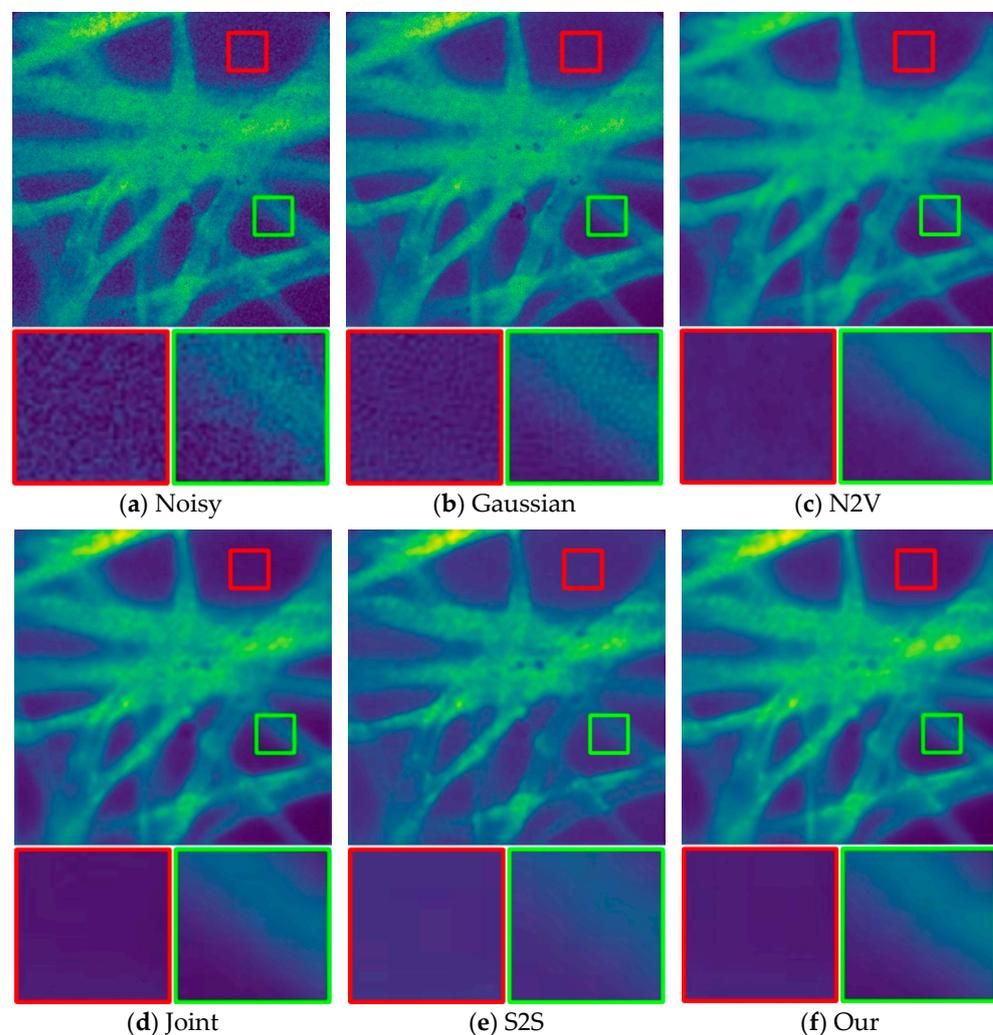


Figure 9. Denoising results of a pCLE image by different methods.

It can be seen that there are still many foveal patterns in the Gaussian denoising method and, compared with other remaining methods, our method is better for the recovered image information, especially in the enlarged green area. For the last four sets of images in Figure 9, two regions of each image are zoomed in, and the red-boxed portion is the background, theoretically devoid of any information. The presence of excessive white

pseudo-shadowing is evident in Figure 9c. The green bordered portion represents the magnification of the sample. It is clear that the contrast in part of Figure 9e is low, making it difficult to discern the location of the optical fiber. Although the contrast difference between Figure 9d,f is not significant, it is shown in subsequent experiments that our method provides more information and details than the JOINT method in “d”.

In order to provide a fair evaluation of image recovery, we also evaluate our method using objective metrics such as PSNR, SSIM, and also compare our method with four self-supervised denoising methods. We downloaded the fluorescence microscopy dataset W2S in the presence of real image and conducted comparison tests with each method on this dataset. Table 1 shows the metrics of all the results, and the PSNR and SSIM of our method are significantly better than the other methods in all three datasets.

Table 1. Evaluation of different methods on three datasets with PSNR and SSIM metric.

Algorithm	W2S-1	W2S-2	W2S-3
	PSNR/SSIM	PSNR/SSIM	PSNR/SSIM
NOISY	17.91/0.3432	15.54/0.2298	14.12/0.2151
GAUSSIAN	30.41/0.5176	30.55/0.4881	31.74/0.4295
N2V	32.73/0.8432	31.24/0.8546	33.50/0.8467
JOINT	32.73/0.8691	31.24/0.8678	33.50/0.8597
S2S	34.98/0.8664	33.22/0.8621	35.64/0.8732
OURS	35.67/0.8794	34.21/0.8721	36.14/0.8984

Through subjective visual assessment, Figures 10–12 demonstrate a notable denoising efficacy of our method on the W2S dataset.

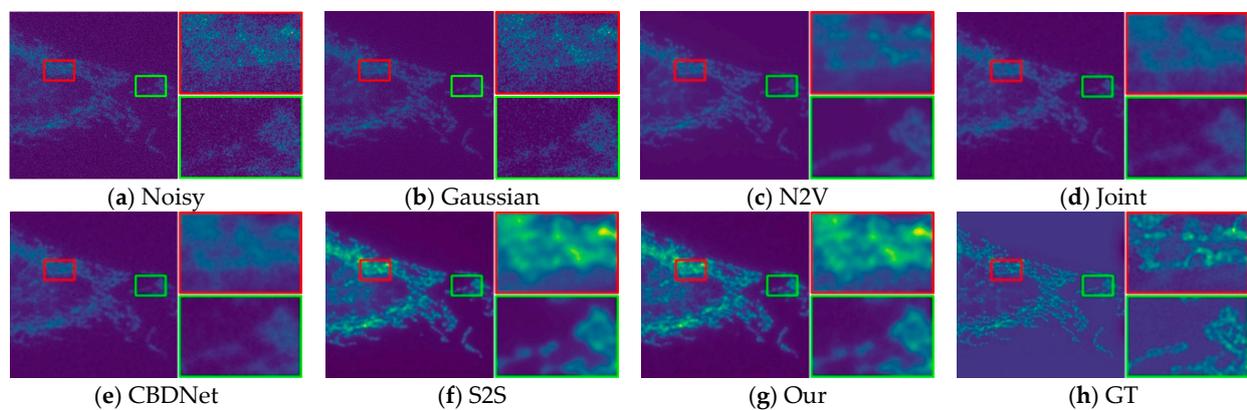


Figure 10. Visual comparison of denoising on the W2S-1 dataset.

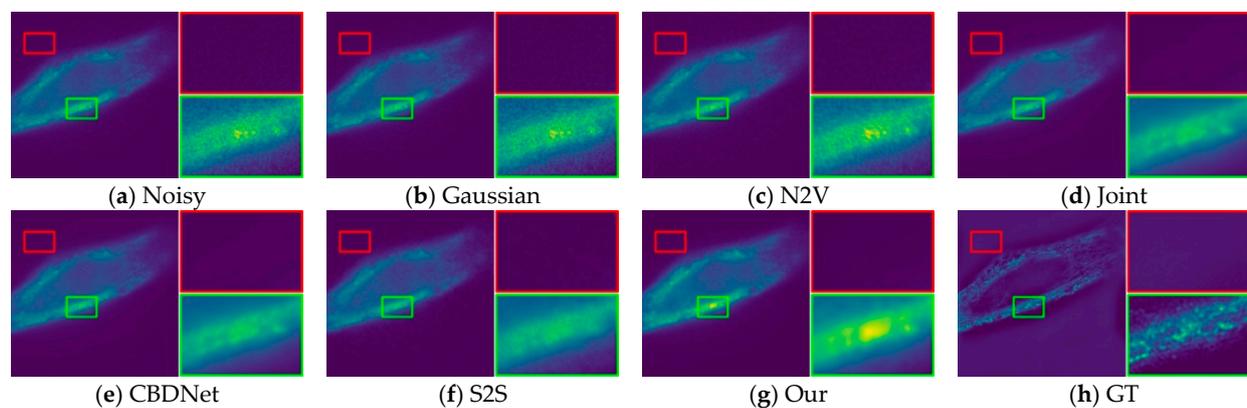


Figure 11. Visual comparison of denoising on the W2S-2 dataset.

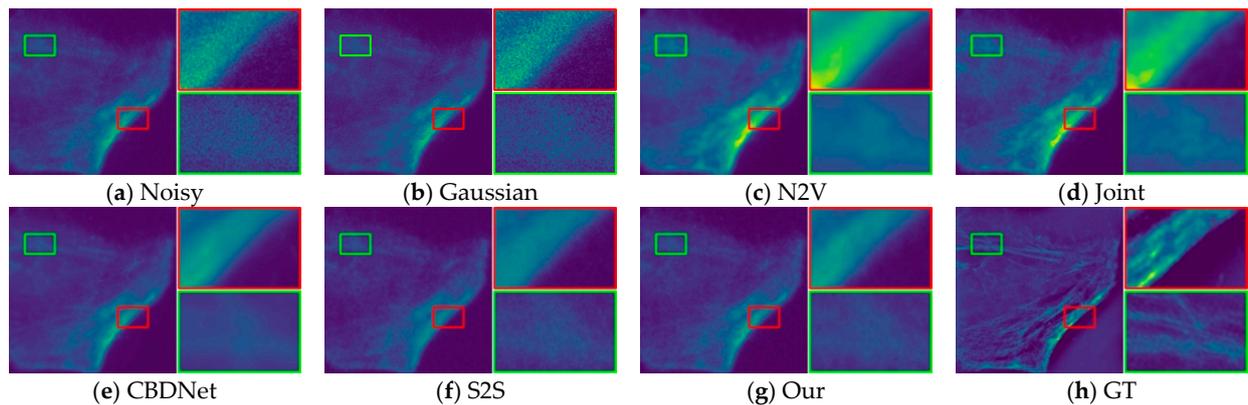


Figure 12. Visual comparison of denoising on the W2S-3 dataset.

These comparisons are important for demonstrating the efficacy of our denoising method and its superiority in terms of objective metrics and subjective visual quality.

The noise prediction network plays a role in the whole algorithmic architecture; it is able to provide more information to make up for the lack of information in D-Net itself and to provide more image features for subsequent training. We first compared the denoising effect of this algorithm with other algorithms, and, in order to verify the role of N-Net, we conducted an ablation experiment, which was used to quantify the role of N-Net. Table 2 demonstrates the comparison of metrics without and with the addition of N-net on the W2S dataset.

Table 2. Ablation study of the contribution of N-Net in terms of PSNR, SSIM.

Discriminator	W2S-1	W2S-2	W2S-3
	PSNR/SSIM	PSNR/SSIM	PSNR/SSIM
N-Net ×	35.64/0.8631	33.79/0.8692	35.38/0.8781
N-Net ✓	35.67/0.8794	34.21/0.8712	36.14/0.8984

It can be seen that N-net extracts effective noise information in training and that this information has a positive effect on the denoising network.

4. Discussion

In this study, we compare the denoising effects of traditional denoising methods, self-supervised denoising methods, and jointly trained denoising methods on laser confocal microscopy imaging. We observe that, in contrast to other self-supervised denoising approaches, the incorporation of additional neural networks for assistance proves more beneficial in effectively extracting informative content from the images. Furthermore, within self-supervised denoising methods, neural networks, leveraging their inherent priors, demonstrate significant denoising capabilities. During the image restoration process, these networks first learn the relevant information within the image before addressing noise, resulting in an initial improvement followed by a deterioration in image quality. To select the most effective images during the recovery process, we employ a quality assessment network that we have trained, facilitating real-time evaluation of image restoration outcomes. On the other hand, we investigate the impact of joint training on the denoising network. Despite prior research in image noise prediction, the generalizability of such methods requires improvement. In this work, we consider the effectiveness of extracting noise information before engaging in joint training, and our results indicate not only superior image quality but also robust performance.

5. Conclusions

In this paper, we propose a self-supervised denoising algorithm to denoise images without clean images such as pCLE images. Current deep learning methods usually cannot be generalized to different samples, but our method achieves good results on fluorescence microscopy images as well. We exploited the structure of neural networks with a priori knowledge to verify that the trend of image recovery is to recover information before noise, and also designed a network to evaluate the results of image recovery by migration learning.

The combination of these three networks can solve the problem of denoising pCLE images and fluorescence microscopy images in complex environments. From the results, it can be seen that our method shows good results in terms of both subjective visual effects and quantitative metrics compared with the current state-of-the-art denoising algorithms.

Author Contributions: Conceptualization, H.Z. and K.Y.; methodology, K.Y., H.Z. and Y.Q.; software, K.Y.; validation, Y.Q., T.Z. and K.Y.; formal analysis, K.Y., H.Z. and Y.Q.; resources, H.Z. and Z.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This work is supported by National Natural Science Foundation of China (62375026) and Fund of the State Key Laboratory of IPOC (BUPT) (IPOC2021ZT06).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data are contained within the article.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Gmitro, A.F.; Aziz, D. Confocal microscopy through a fiber-optic imaging bundle. *Opt. Lett.* **1993**, *18*, 565–567. [[CrossRef](#)] [[PubMed](#)]
2. Hu, B.; Bolus, D.; Brown, J.Q. Improved contrast in inverted selective plane illumination microscopy of thick tissues using confocal detection and structured illumination. *Biomed. Opt. Express* **2017**, *8*, 5546–5559. [[CrossRef](#)] [[PubMed](#)]
3. Hughes, M.; Yang, G.-Z. Line-scanning fiber bundle endomicroscopy with a virtual detector slit. *Biomed. Opt. Express* **2016**, *7*, 2257–2268. [[CrossRef](#)]
4. Thompson, A.J.; Hughes, M.; Anastasova, S.; Conklin, L.S.; Thomas, T.; Leggett, C.; Faubion, W.A.; Miller, T.J.; Delaney, P.; Lacombe, F.; et al. The potential role of optical biopsy in the study and diagnosis of environmental enteric dysfunction. *Nat. Rev. Gastroenterol. Hepatol.* **2017**, *14*, 727–738. [[CrossRef](#)] [[PubMed](#)]
5. Thrapp, A.D.; Hughes, M.R. Reduced motion artifacts and speed improvements in enhanced line-scanning fiber bundle endomicroscopy. *J. Biomed. Opt.* **2021**, *26*, 056501. [[CrossRef](#)] [[PubMed](#)]
6. Hughes, M.; Yang, G.-Z. High speed, line-scanning, fiber bundle fluorescence confocal endomicroscopy for improved mosaicking. *Biomed. Opt. Express* **2015**, *6*, 1241–1252. [[CrossRef](#)]
7. Olivas, S.J.; Arianpour, A.; Stamenov, I.; Morrison, R.; Stack, R.A.; Johnson, A.R.; Agurok, I.P.; Ford, J.E. Image processing for cameras with fiber bundle image relay. *Appl. Opt.* **2015**, *54*, 1124–1137. [[CrossRef](#)] [[PubMed](#)]
8. Yao, B.; Huang, B.; Li, X.; Qi, J.; Li, Y.; Shao, Y.; Qu, J.; Gu, Y.; Li, J. Depixelation and image restoration with meta-learning in fiber-bundle-based endomicroscopy. *Opt. Express* **2022**, *30*, 5038–5050. [[CrossRef](#)]
9. Lee, C.-Y.; Han, J.-H. Elimination of honeycomb patterns in fiber bundle imaging by a superimposition method. *Opt. Lett.* **2013**, *38*, 2023–2025. [[CrossRef](#)]
10. Broadus, C.; Krull, A.; Weigert, M.; Schmidt, U.; Myers, G. Removing structured noise with self-supervised blind-spot networks. In Proceedings of the 2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI), Iowa City, IA, USA, 3–7 April 2020; pp. 159–163.
11. Liu, X.; Zhang, L.; Kirby, M.; Becker, R.; Qi, S.; Zhao, F. Iterative l1-min algorithm for fixed pattern noise removal in fiber-bundle-based endoscopic imaging. *J. Opt. Soc. Am. A* **2016**, *33*, 630–636. [[CrossRef](#)]
12. Perperidis, A.; Dhaliwal, K.; McLaughlin, S.; Vercauteren, T. Image computing for fibre-bundle endomicroscopy: A review. *Med. Image Anal.* **2020**, *62*, 101620. [[CrossRef](#)] [[PubMed](#)]
13. Tian, L.; Hunt, B.; Bell, M.A.L.; Yi, J.; Smith, J.T.; Ochoa, M.; Intes, X.; Durr, N.J. Deep Learning in Biomedical Optics. *Lasers Surg. Med.* **2021**, *53*, 748–775. [[CrossRef](#)] [[PubMed](#)]
14. Liu, J.; Zhou, W.; Xu, B.; Yang, X.; Xiong, D. Honeycomb pattern removal for fiber bundle endomicroscopy based on a two-step iterative shrinkage thresholding algorithm. *AIP Adv.* **2020**, *10*, 045004. [[CrossRef](#)]

15. Weigert, M.; Schmidt, U.; Boothe, T.; Müller, A.; Dibrov, A.; Jain, A.; Wilhelm, B.; Schmidt, D.; Broaddus, C.; Culley, S.; et al. Content-aware image restoration: Pushing the limits of fluorescence microscopy. *Nat. Methods* **2018**, *15*, 1090–1097. [[CrossRef](#)] [[PubMed](#)]
16. Lecun, Y.; Bottou, L.; Bengio, Y.; Haffner, P. Gradient-based learning applied to document recognition. *Proc. IEEE* **1998**, *86*, 2278–2324. [[CrossRef](#)]
17. LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436–444. [[CrossRef](#)] [[PubMed](#)]
18. Zhao, H.; Ni, B.; Liu, W.; Jin, X.; Zhang, H.; Gao, X.W.; Wen, X.; Shi, D.; Dong, L.; Xiong, J.; et al. Signal denoising of viral particle in wide-field photon scattering parametric images using deep learning. *Opt. Commun.* **2022**, *503*, 127463. [[CrossRef](#)]
19. Luisier, F.; Blu, T.; Unser, M. Image denoising in mixed poisson–gaussian noise. *IEEE Trans. Image Process.* **2011**, *20*, 696–708. [[CrossRef](#)] [[PubMed](#)]
20. Vyas, K. High-Resolution Fluorescence Endomicroscopy for Rapid Evaluation of Breast Cancer Margins. Ph.D. Thesis, Imperial College London, London, UK, 2018.
21. Wang, Y.; Pinkard, H.; Khwaja, E.; Zhou, S.; Waller, L.; Huang, B. Image denoising for fluorescence microscopy by supervised to self-supervised transfer learning. *Opt. Express* **2021**, *29*, 41303–41312. [[CrossRef](#)]
22. Plotz, T.; Roth, S. Benchmarking denoising algorithms with real photographs. In Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 June 2017.
23. Lebrun, M.; Colom, M.; Morel, J.-M. The noise clinic: A blind image denoising algorithm. *Image Process. Line* **2015**, *5*, 1–54. [[CrossRef](#)]
24. Zhang, K.; Li, Y.; Liang, J.; Cao, J.; Zhang, Y.; Tang, H.; Fan, D.-P.; Timofte, R.; Van Gool, L. Practical Blind Image Denoising via Swin-Conv-UNet and Data Synthesis. *Mach. Intell. Res.* **2023**, *20*, 822–836. [[CrossRef](#)]
25. Wang, H.; Rivenson, Y.; Jin, Y.; Wei, Z.; Gao, R.; Günaydin, H.; Bentolila, L.A.; Kural, C.; Ozcan, A. Deep learning enables cross-modality super-resolution in fluorescence microscopy. *Nat. Methods* **2019**, *16*, 103–110. [[CrossRef](#)] [[PubMed](#)]
26. Goncharova, A.S.; Honigsmann, A.; Jug, F.; Krull, A. Improving blind spot denoising for microscopy. In *European Conference on Computer Vision*; Springer: Berlin/Heidelberg, Germany, 2020; pp. 380–393.
27. Meister, A. On the effect of misspecifying the error density in a deconvolution problem. *Can. J. Stat.* **2004**, *32*, 439–449. [[CrossRef](#)]
28. Lehtinen, J.; Munkberg, J.; Hasselgren, J.; Laine, S.; Karras, T.; Aittala, M.; Aila, T. Noise2noise: Learning image restoration without clean data. *arXiv* **2018**, arXiv:1803.04189.
29. Krull, A.; Buchholz, T.-O.; Jug, F. Noise2void-learning denoising from single noisy images. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 2129–2137.
30. Batson, J.; Royer, L. Noise2self: Blind denoising by self-supervision. In Proceedings of the International Conference on Machine Learning, Long Beach, CA, USA, 10–15 June 2019; pp. 524–533.
31. Ollion, J.; Ollion, C.; Gassiat, E.; Lehericy, L.; Corff, S.L. Joint self-supervised blind denoising and noise estimation. *arXiv* **2021**, arXiv:2102.08023.
32. Ulyanov, D.; Vedaldi, A.; Lempitsky, V. Deep image prior. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, 18–23 June 2018.
33. Quan, Y.; Chen, M.; Pang, T.; Ji, H. Self2self with dropout: Learning self-supervised denoising from single image. In Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 13–19 June 2020.
34. Tang, Y.; Yang, D.; Li, W.; Roth, H.R.; Landman, B.; Xu, D.; Nath, V.; Hatamizadeh, A. Self-supervised pre-training of swin transformers for 3d medical image analysis. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 18–24 June 2022; pp. 20730–20740.
35. Zou, S.; Long, M.; Wang, X.; Xie, X.; Li, G.; Wang, Z. A cnn-based blind denoising method for endoscopic images. In Proceedings of the 2019 IEEE Biomedical Circuits and Systems Conference (BioCAS), Nara, Japan, 17–19 October 2019; pp. 1–4.
36. Guo, S.; Yan, Z.; Zhang, K.; Zuo, W.; Zhang, L. Toward convolutional blind denoising of real photographs. In Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 15–20 June 2019; pp. 1712–1722.
37. Gassiat, E.; Le Corff, S.; Lehe, L. Deconvolution with unknown noise distribution is possible for multivariate signals. *Ann. Stat.* **2022**, *50*, 303–323. [[CrossRef](#)]
38. Tian, X.; Wu, Q.; Wei, H.; Zhang, Y. Noise2sr: Learning to denoise from super-resolved single noisy fluorescence image. In Proceedings of the Medical Image Computing and Computer Assisted Intervention—MICCAI 2022: 25th International Conference, Singapore, 18–22 September 2022; pp. 334–343.
39. Zhou, R.; El Helou, M.; Sage, D.; Laroche, T.; Seitz, A.; Süsstrunk, S. W2s: Microscopy data with joint denoising and super-resolution for widefield to sim mapping. In Proceedings of the European Conference on Computer Vision, Glasgow, UK, 23–28 August 2020; pp. 474–491.
40. Ronneberger, O.; Fischer, P.; Brox, T. U-net: Convolutional networks for biomedical image segmentation. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Munich, Germany, 5–9 October 2015; pp. 234–241.
41. Su, S.; Yan, Q.; Zhu, Y.; Zhang, C.; Ge, X.; Sun, J.; Zhang, Y. Blindly assess image quality in the wild guided by a self-adaptive hyper network. In Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 13–19 June 2020.

42. Wang, Z.; Bovik, A.C.; Sheikh, H.R.; Simoncelli, E.P. Image quality assessment: From error visibility to structural similarity. *IEEE Trans. Image Process.* **2004**, *13*, 600–612. [[CrossRef](#)]
43. Ma, J.; Li, S.; He, J.; Wang, Y.; Liao, Y.; Zeng, D.; Bian, Z. Blind CT image quality assessment via deep learning strategy: Initial study. In *Medical Imaging 2018: Image Perception, Observer Performance, and Technology Assessment*; Nishikawa, R.M., Samuelson, F.W., Eds.; International Society for Optics and Photonics (SPIE): San Diego, CA, USA, 2018; Volume 10577, p. 105771A.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.