photonics

MDPI

Article

High-Resolution Image Processing of Probe-Based Confocal Laser Endomicroscopy Based on Multistage Neural Networks and Cross-Channel Attention Module

Yufei Qiu [1], Haojie Zhang [1,*], Kun Yang [1], Tong Zhai [1], Yipeng Lu [2], Zhongwei Cao [1] and Zhiguo Zhang [1]

[1] State Key Lab of Information Photonics and Optical Communications, Beijing University of Posts and Telecommunications (BUPT), Beijing 100876, China; qiuyufei@bupt.edu.cn (Y.Q.); ykk@bupt.edu.cn (K.Y.); zhai_tong@bupt.edu.cn (T.Z.); caozhongwei@bupt.edu.cn (Z.C.); zhangzhiguo@bupt.edu.cn (Z.Z.)
[2] School of Integrated Circuit Science and Engineering, Beihang University, Beijing 100876, China; luyipeng@buaa.edu.cn
* Correspondence: haojie.zhang@bupt.edu.cn

Abstract: Probe-based confocal laser endomicroscopy (pCLE) is a subcellular in vivo imaging technique that generates diagnostic images revealing malignant structural modifications in epithelial tissues. In the clinical diagnosis of probe confocal laser endomicroscopy (pCLE), the image background generally has the problems of dynamic blur or information loss, which is not conducive to achieving high-resolution and clear pCLE imaging. In recent years, deep learning technology has achieved remarkable results in image deblurring. For the task of recovering high-resolution pCLE images, the current methods still suffer from the following drawbacks: it is difficult to choose a strategy to make CNN converge at a deeper level and mainstream methods cannot handle the complex balance between spatial details and high-level feature information well when reconstructing clear images. In order to solve the problem, we propose a new cross-channel attention, multistage, high-resolution pCLE image deblurring structure. This methodology improves the supervised attention mechanism, enhances the ability of feature extraction and fusion capabilities, and improves the quality of image deblurring by adding cross-channel attention module (CAM) into the multistage neural networks' architecture. The experimental results show that the average peak signal-to-noise ratio (PSNR) of the proposed model on the dataset is as high as 29.643 dB, and the structural similarity (SSIM) reaches 0.855. This method is superior to the prior algorithms in the visualization of recovered images, and the edge and texture details of the restored pCLE images are clearer.

Keywords: probe confocal laser endomicroscopy; cross-channel attention module; multistage neural networks; image deblurring

Citation: Qiu, Y.; Zhang, H.; Yang, K.; Zhai, T.; Lu, Y.; Cao, Z.; Zhang, Z. High-Resolution Image Processing of Probe-Based Confocal Laser Endomicroscopy Based on Multistage Neural Networks and Cross-Channel Attention Module. *Photonics* **2024**, *11*, 106. https://doi.org/10.3390/photonics11020106

Received: 21 December 2023
Revised: 11 January 2024
Accepted: 15 January 2024
Published: 25 January 2024

1. Introduction

Probe-based confocal laser endomicroscopy (pCLE) is a micro-endoscopic imaging diagnostic technique widely used in the medical field in recent years and has achieved considerable clinical value [1]. In the field of endoscopic microscopy, commonly known as "endomicroscopy", coherent fiber bundle image guides offer a practical solution for transmitting images from the tissue to external microscope optics [1,2]. This proves especially beneficial for confocal imaging, as it allows the laser scanning system to be positioned outside the patient [2]. The systems rely on fiber optics, employ laser technology, and necessitate rapid and accurate scanning mechanisms [3].

In the process of collecting the existing pCLE images, optical imaging systems face inherent limitations due to the physical characteristics of their components [4], encompassing factors such as the lens's shape and material, as well as the shooting location. These constraints give rise to phenomena like scattering and refraction as light traverses the lens or reflector. In real-time image acquisition, vision sensors may introduce blurring,

leading to changes in image resolution and contrast [5,6]. Therefore, the image background generally has the problems of dynamic blur or information loss, which bring great troubles to the realization of high-resolution and clear pCLE image [5], and seriously affects the analysis and follow-up research of image information. Consequently, the nature of image acquisition through the fiber bundle gives rise to several inherent characteristics and limitations necessitating novel and effective image preprocessing and postprocessing algorithms, ranging from image formation and enhancement to pathology detection and quantification [5]. The task of image deblurring holds significant importance in the realm of optical imaging, playing a crucial role in enhancing overall image quality [7].

Deep convolutional neural networks (CNNs) are a new type of network and have the ability to automatically effectively learn the nonlinear relationship between input and output [8–10]. Recent studies have shown that deep convolutional neural networks (CNNs) are widely used in image restoration tasks [11–15] in computer vision. For instance, Liu et al. [16] proposed a novel approach that merged multilevel image restoration with the pix2pix generative adversarial network architecture within the lensless imaging sphere, which greatly improved image recovery quality in lensless systems. Zhang et al. [4] proposed a multiscale circular image deblurring model based on PVC-Resnet in order to achieve the restoration of different scale objects in blurred images and obtain the global features of blurred images. Cheng et al. [17] designed a method to mitigate atmospheric turbulence using optical flow and convolutional neural networks, thus reducing the turbulence mitigation problem to a deblurring problem. Asim et al. [18] regularized the ill-posed and nonlinear blind image deconvolution using deep generative networks as priors, which is able to achieve good deblurring of images with large blurs and heavy measurement noise.

For the task of recovering high-resolution pCLE images, in order to enable the CNN to converge at a deeper level [19], as well as to better handle the complex balance between spatial details and high-level feature information when reconstructing clear images [19], we propose a high-resolution pCLE image deblurring method based on cross-channel attention module (CAM) enhancement. Our proposed approach consists of multiple stages, and each stage focuses on refining the restored image using the CAM [20]. In the initial stage, the low-resolution input image is upscaled using conventional techniques [21]. Subsequent stages employ the CAM to aggregate multiscale deep textures for capturing finer details and preserving structural information [10]. The model, trained on mixed-size patches via progressive learning, shows enhanced performance at test time, where images can be of different resolutions [20]. This progressive refinement process can improve the quality of image restoration and produce high-resolution images with enhanced visual fidelity [22,23]. The experimental results indicate that the proposed method outperforms the comparative algorithms in both quantitative and qualitative analyses on synthetic and real datasets. The restored images retain more clear details, which is particularly important in applications such as medical imaging and surveillance.

2. Method

2.1. Cross-Channel Attention Module

To avoid the time and memory complexity of the key-query dot product interaction grows quadratically with the spatial resolution of input, we propose CAM, shown in Figure 1. High-resolution images can be efficiently processed by aggregating local and nonlocal pixel interactions [23]. Another important aspect of CAM is its linear complexity. The key factor is to apply SA across channels. We first design a deep convolution module in CAM to emphasize the local context. We calculate the cross covariance between channels to generate an attention map that contains the global context.

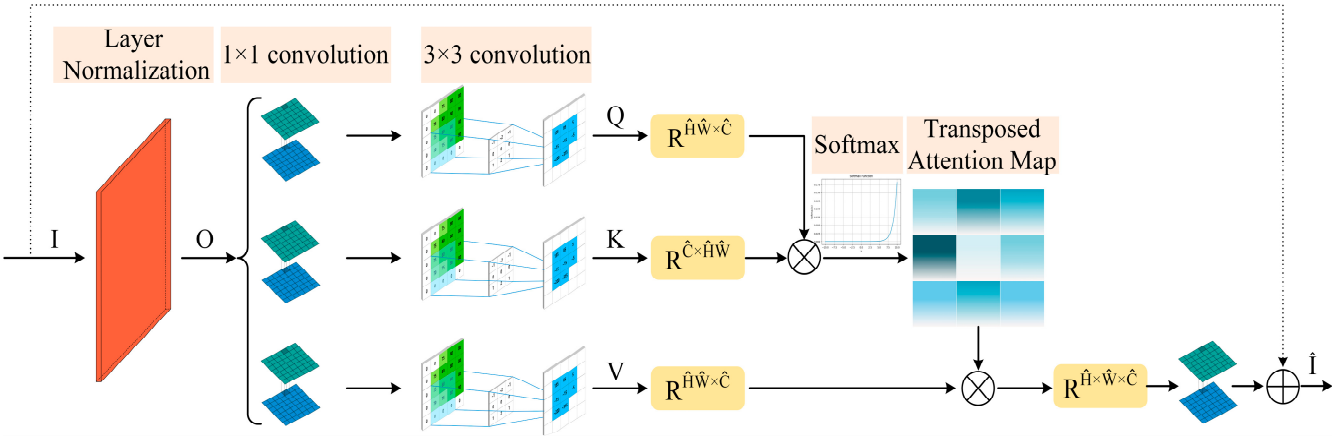


Figure 1. The processing architecture of proposed CAM.

Based on the layer normalization tensor $O \in R^{\hat{H}\times\hat{W}\times\hat{C}}$, our CAM can obtain query (Q), key (K), and value (V) projections, combined with local context information. Our CAM applies 1×1 convolution and 3×3 deep convolutions to aggregate per-pixel cross-channel context and encode per-channel spatial context, respectively. The calculation formula for the result is as in Equations (1)–(3), where $W_i^{(\cdot)}$ is the 1×1 per-pixel convolution and $W_j^{(\cdot)}$ is the 3×3 depth-wise convolutions. Convolutional layers in the network are unbiased, which also reduces the computation.

$$Q = W_i^Q W_j^Q O \tag{1}$$

$$K = W_i^K W_j^K O \tag{2}$$

$$V = W_i^V W_j^V O \tag{3}$$

The key next step is to generate the transposed attention map A by interacting the newly generated query with the key projections dot product. Compared to the huge size of the traditional attention map, the size of the transposed attention map A is $R^{\hat{C}\times\hat{C}}$. Overall, the computation process of CAM is defined as in Equation (4):

$$\hat{I} = W_j \hat{V} \cdot softmax\left(\hat{K} \cdot \hat{Q} / \beta\right) + I \tag{4}$$

where I and $\hat{I}$ are the input and output feature maps; next, the newly generated matrices based on the original tensor are $\hat{Q} \in R^{\hat{H}\hat{W}\times\hat{C}}$, $\hat{K} \in R^{\hat{C}\times\hat{H}\hat{W}}$, $\hat{V} \in R^{\hat{H}\hat{W}\times\hat{C}}$. At the same time, we set the learnable scaling parameter β before the softmax function to control the dot product of $\hat{K}$ and $\hat{Q}$.

2.2. Multistage Feature Processing and Fusion

Breaking through the limitation of traditional single-stage CNN in image deblurring task, there are two ideal models in network, the encoder–decoder model and the single-scale feature model [21]. The advantage of the encoder–decoder model is that it can capture multiscale information efficiently. Deficiency is the limitation of spatial detail loss due to repeated downsampling. The advantage of the single-scale model is the ability to preserve spatial detail. However, the disadvantage is that the vision is limited and the global depth information cannot be grasped. We propose a hybrid multistage model that combines the advantages of both approaches. The encoder–decoder model is applied in the initial phase of the network, and the network with the original input resolution is used in the final phase.

Firstly, we use the attention module to collect the features at each scale, and also process the feature mapping of U-Net skip connection. The bilinear upsampling followed

by the convolution layer helps to reduce checkerboard artifacts that often occur in the output image. Encoder–decoder model design is based on standard U-net. Next, we apply the single-scale model in order to preserve fine details from the input image to the output image. This part does not use downsampling and can generate high-resolution features with rich information. This part consists of multiple raw resolution blocks, each of which contains channel attention information. Finally, features need to be refined both within and between encoder–decoder models. In our method, 1×1 convolution refinement is required before the features from the previous stage are transferred to the next stage.

The proposed feature processing and fusion method has several merits. By reducing information loss in repeated scaling operations, enriching the multiscale characteristics of subsequent phases, and promoting stable network optimization, more phases are allowed to be added to the architecture [24,25].

2.3. Core Network Architecture

The overall network architecture consists of three phases. The first two stages mainly apply the encoder–decoder model, meaning that the context information can be fully learned with a large field of view. The third stage applies a model that operates on the resolution of the original input image (without any downsampling operations), satisfying the pixel-to-pixel correspondence from input to output, and ultimately preserving fine spatial details in the output image.

Instead of simply stacking three phases on top of each other, the network architecture adds a truth value attention module to form the filter (AM filter) in front of the CAM at each phase. Instead of predicting images directly, it uses the supervision of real tags to guide the image recovery more effectively. AM filter can selectively filter and transmit only relevant features to subsequent stages by generating attention maps. This process involves generating a residual image R_S from the incoming feature and then combining it with the input image I to produce a restored image Q_S. Based on Q_S, attention masks for each pixel can be obtained, and Q_S is directly supervised by the real image. These masks recalibrate local features to create an enhanced feature set for subsequent processing. The loss function of the overall network is defined as in Equations (5)–(8):

$$Q_S = I + R_S \tag{5}$$

$$\mathcal{L} = \sum_{S=1}^{3} [\mathcal{L}_1 + \lambda \mathcal{L}_2] \tag{6}$$

$$\mathcal{L}_1 = \sqrt{||Q_S - T|| + \varepsilon^2} \tag{7}$$

$$\mathcal{L}_2 = \sqrt{||\Delta(Q_S) - \Delta(T)||^2 + \varepsilon^2} \tag{8}$$

where $\mathcal{L}_1$ is the Charbonnier loss, and T represents the ground-truth image. ε is commonly used as a small positive constant in various optimization algorithms, especially in scenarios involving numerical computations. Choosing a smaller ε value helps prevent division by zero issues during gradient calculations or related computations. The constant ε is set to 10^{-3} in the experiment. $\mathcal{L}_2$ is the edge loss, where Δ denotes the Laplacian operator. λ represents the mixing coefficients generated through the Beta distribution. The Beta distribution is commonly used to generate random numbers between 0 and 1, which aligns well with the range of mixing coefficients. In addition, parameter λ is set to control the relative proportion of the two losses to achieve the optimal effect.

The proposed core network architecture for pCLE deblurring, shown in Figure 2. Each stage has the application of the original input image. The difference is that the image is divided into different patches in different stages: four patches for Stage 1, two patches for Stage 2 and the original image for the last stage [21]. The two attention modules (CAM and AM filter) have different functions to achieve better feature extraction and processing effects. This is also a prominent innovation point of pCLE deblurring architecture design.

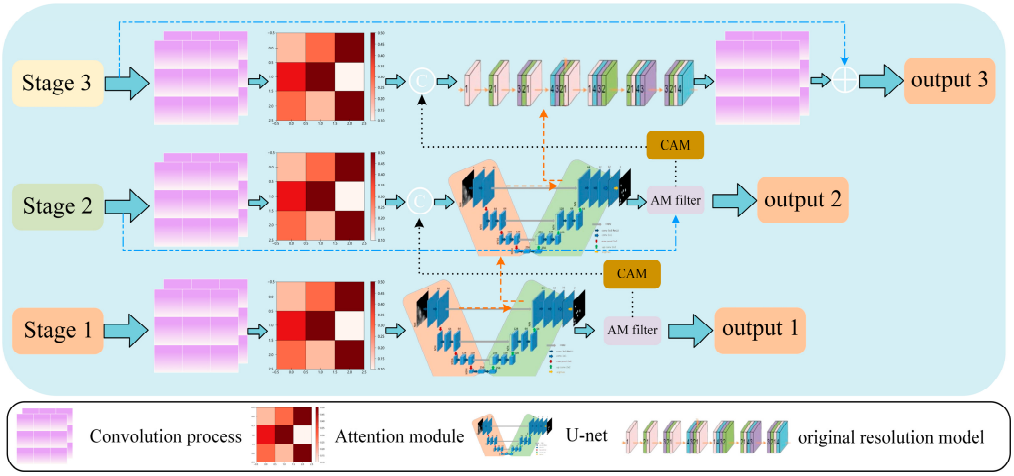


Figure 2. The core network architecture for pCLE deblurring. The first two stages mainly apply the encoder–decoder mode. The third stage applies a model that operates on the resolution of the original input image.

3. Experiment

In this section, we propose the network model and the CAM-enhanced high-resolution pCLE deblurring method and evaluate the optimization results of this method quantitatively and qualitatively.

3.1. Dataset

Due to the absence of ground truth in pCLE images, we constructed a hybrid training dataset. Part of it is from the GoPro dataset [26] and the other part is obtained from the existing FMD dataset [27] acquired under the fluorescence microscopy. Among them, since the original FMD dataset image is mainly used for denoising, we have carried out innovation on the dataset. The input images of the FMD dataset are partially constructed by adding several existing major ambiguities (Gaussian blur, bilateral filter) that need to be overcome. The overall training dataset comprises 2903 input images. The GoPro dataset contains 2103 images and the FMD dataset contains 800 images. To verify the applicability of the method, we apply our new training model directly to the Widefield2SIM (W2S) dataset [28]. Similarly, the input image for W2S is added with the same ambiguity as the input for the training dataset. It's crucial to note that the training and test datasets are independent and distinct. In this way, the deblurring effect of our proposed method can be evaluated accurately and objectively. In addition to the intuitive visual evaluation, quantitative comparisons are performed using the PSNR (peak signal-to-noise ratio) and SSIM (structural similarity index measure) [29] metrics.

3.2. Implementation Details

The model is built using PyTorch 2.0.0 and tested using multiple GPUs in the environment with Python 3.9.13 and CUDA 11.7. Our approach does not require pretraining, it is end to end and can be trained directly. We train the model for the pCLE deblurring task. Since CAM is capable of modeling global connectivity and remains applicable to large images, we apply CAM between the first and second stages, as well as between the second and third stages. CAM employs 1×1 convolution for pixel-wise aggregation of cross-channel context and utilizes efficient deep convolutions for channel-wise aggregation of local context [20]. This strategy emphasizes spatial local context and ensures that the contextualized global relationships between pixels are implicitly modeled [20]. We use 2×2 max-pooling with stride 2 for downsampling. In the third stage, we use a raw resolution network, each of which further applies multiple raw resolution processing blocks for continuous feature acquisition of the input information. For different levels of processing

tasks, we can change the number of channels to affect the width of the network. For example, in our application, when deblurring the pCLE images, the number of channels is set to 96. To augment the data, we use random horizontal and vertical flips. The networks are trained on 256 × 256 patches with a batch size of 4 for 4×10^5 iterations. We use the Adam optimizer [30] and set the initial learning rate to 2×10^{-4}, which is steadily decreased to 1×10^{-6} through the cosine annealing strategy [21].

3.3. Evaluation on the Image Deblurring Network

In order to ensure that the model is well trained while avoiding overfitting, we have implemented several key measures and metrics. Throughout the entire training process, we continuously monitor relevant parameter indicators, such as the loss value, training duration, and changes in learning rate, along with other performance metrics. We dynamically adjust the learning rate during training to prevent overshooting and enhance convergence. By tracking these metrics, if the performance on the training set ceases to improve or starts to decline, we can halt the training process. This helps prevent overfitting, ensuring that the model does not become too closely aligned with the training data. We have conducted multiple repeated training experiments, striking a balance between fitting the training data and achieving a high level of generalization. Ultimately, this ensures that our model is well trained and capable of effectively handling unseen data. During the training process, we record the PSNR value of the model after each epoch round, where the best model is defined as the generated model when the best PSNR value is reached. The following Figure 3 provides a detailed illustration of the variation in the PSNR value of the model with increasing epochs during the training process. It is obvious that the PSNR value in the early stage presents an overall increasing trend, and reaches the best value at the 74th epoch, when the value is 22.2603 dB, and the best training model is obtained at this time. Subsequently, as the training epoch increases, after the 80th epoch, the PSNR value will show a stable and convergent state, so that the point of maximum value represents the optimal model. We obtain the best model and it can be used in our image deblurring test.

In this work, we evaluate the deblurring method on a real-world dataset. The objective is to assess the effectiveness of the proposed method for deblurring real-world images. We download the fluorescence microscopy dataset W2S. Here we directly use the best model generated by the training to test. The W2S dataset has ground-truth images and is therefore highly representative and persuasive in evaluating our proposed method. We evaluate 120 images from the W2S dataset to ensure a balanced representation of different imaging conditions, such as sample types and imaging methods. As is intuitively shown in the figure below, Figure 4 shows the partially deblurring W2S images resulting from the evaluated method. Figure 4 shows three processing effect diagrams. In each row, from left to right, there are input blurring images, optimization result images and ground truth images. Visual evaluation shows that the images recovered by our model are clearer and closer to reality than other models. From the perspective of objective quantitative indicators, we calculated the PSNR value and SSIM value of the 120 recovered images. The PSNR value is calculated using standardized formula and compared with the ground truth image. The SSIM value is commonly employed as a standard measure in the evaluation of image restoration techniques. We represent the results as a line chart in Figure 5. The PSNR value is represented by a blue line with circular markers. A higher PSNR indicates better image quality. In our dataset, PSNR values exhibit significant fluctuations. However, during specific epochs (such as the 24th, 65th, and 67th epochs), PSNR values reach notably high levels. The overall average PSNR value for the dataset is 29.643 dB. The SSIM value is depicted by a red line with "x" shaped markers, ranging from 0 to 1. Higher values indicate greater similarity between images. Similarly, SSIM values in our dataset show fluctuations, with certain epochs (like the 24th, 65th, and 67th epochs) displaying particularly high SSIM values. The average SSIM value for the entire dataset is 0.855. Quantitative analysis demonstrates that our method effectively improves the overall deblurring of W2S images.

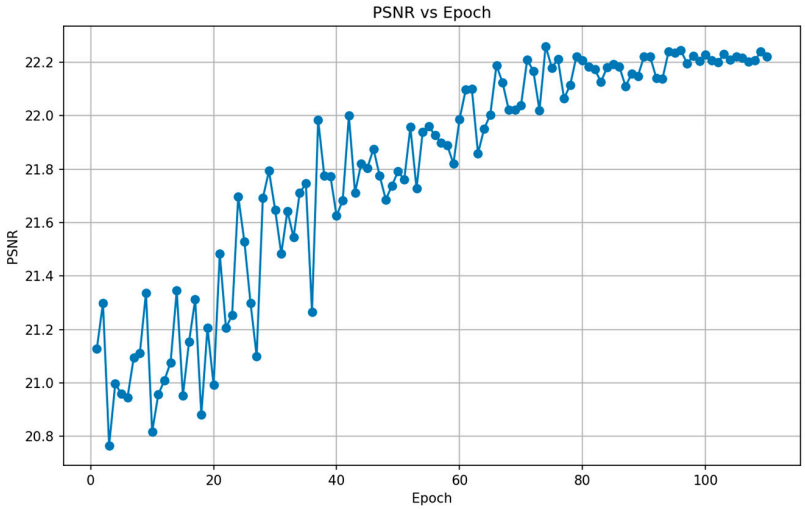
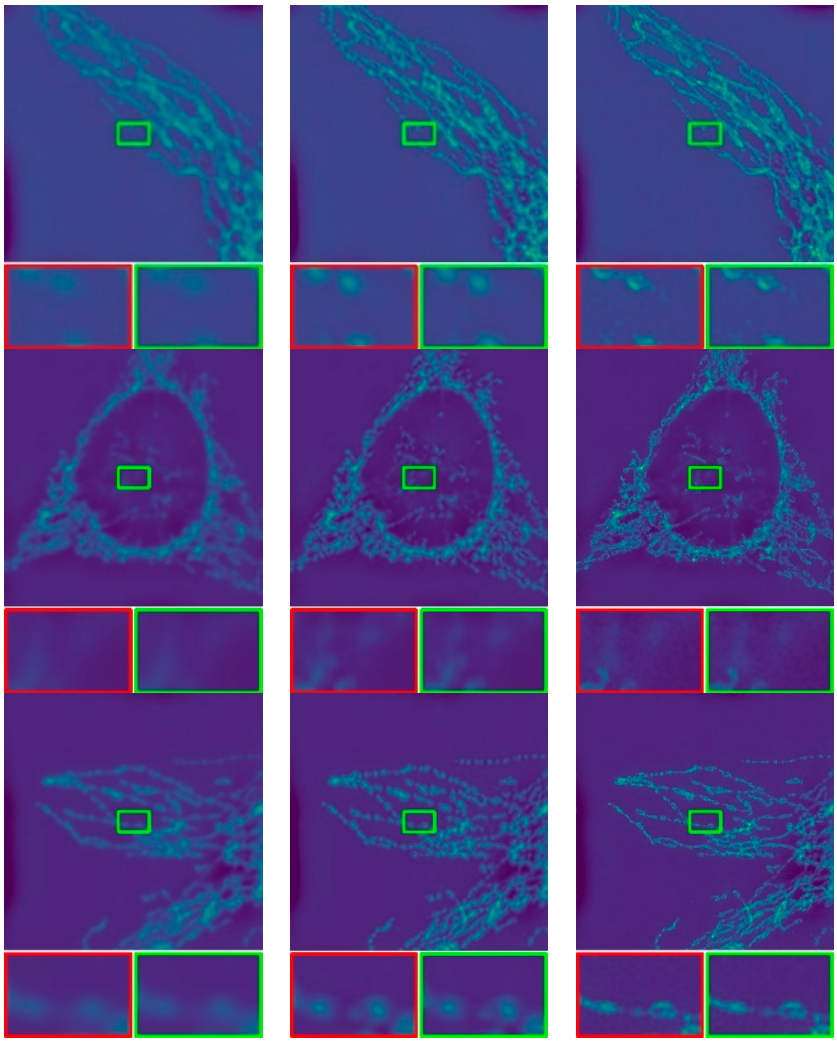


Figure 3. The change curve of the PSNR value of the model with the training epoch.



(**a**) Blurring image (**b**) Ours (**c**) Ground truth

Figure 4. Deblurring results of the W2S: (**a**) the blurring image of the W2S; (**b**) the results of our method; (**c**) the ground truth of the W2S. The red and green framed portions are both locally enlarged images, allowing for a more intuitive comparison of image clarity by magnifying specific details.

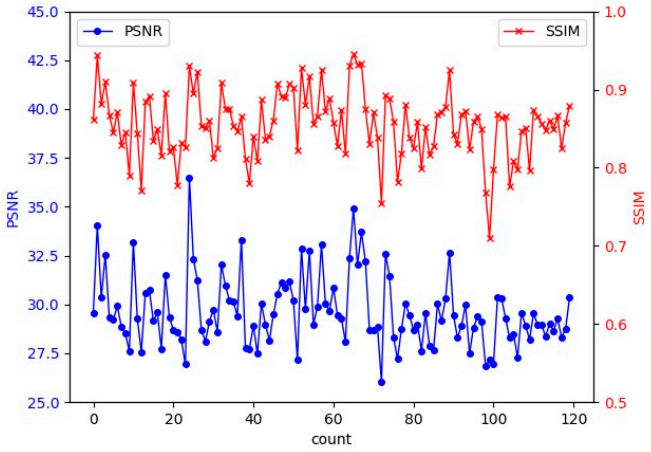
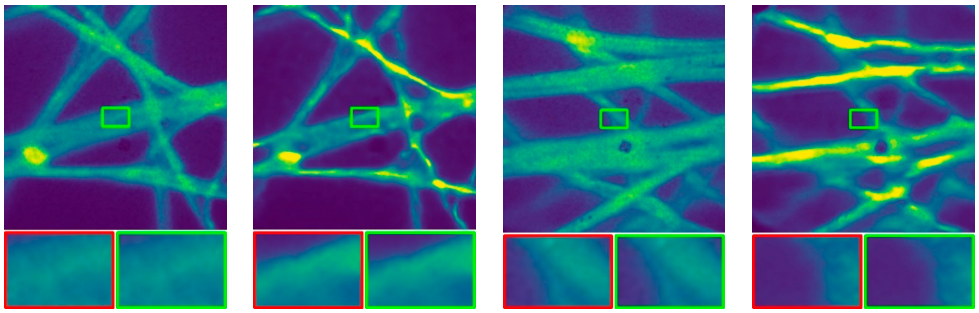


Figure 5. Line plot of the PSNR/SSIM scores for 120 restored images.

3.4. Evaluate the Deblurring Effect on Real pCLE Images

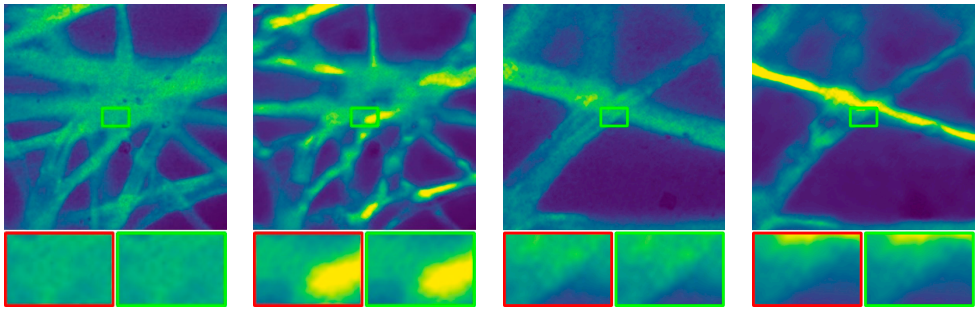
The significance of this performance gain extends beyond numerical metrics; it underscores the practical applicability of our proposed pCLE image deblurring method in real-world scenarios. The recovered images not only exhibit higher fidelity but also closely approximate the visual quality of the original scenes. This is a crucial aspect, especially in applications where image clarity and fidelity are paramount, such as in medical imaging or surveillance. Based on this, we use confocal laser endomicroscopy (pCLE) to capture real biological images and test the deblurring effect of our proposed method. Our pCLE imaging samples are fluorescently dyed tissue paper. We actually evaluate the deblurring effect of 54 pCLE images. Our pCLE images are obtained through real-time imaging with fiber-bundle endomicroscopy. Due to various physical conditions, absolute ground-truth images are not available for these images. Since there is no ground truth, we visually demonstrate the deblurring results obtained by our method through the pCLE deblurring images in Figure 6. Figure 6 shows the deblurring results of four groups of pCLE images. The left and right sides of each group are the acquired input images and the optimized result images, respectively. In each set of optimized result images, clear restoration of crucial highlighted areas is visible. Moreover, our proposed method proficiently alleviates the overall blurring of the images, showcasing its effectiveness in enhancing visual clarity. It can be seen that our method can really achieve a certain degree of deblurring effect, which is of great significance.

By addressing the challenges associated with image deblurring in both synthetic and real-world contexts, our method stands out as a robust and adaptable solution. This bodes well for its potential deployment in a variety of fields where image clarity is of utmost importance, offering a reliable tool for improving the fidelity of visual information in diverse settings.



(**a**) Input and result (**b**) Input and result

Figure 6. *Cont.*



(**c**) Input and result (**d**) Input and result

Figure 6. The deblurring results of four groups of pCLE images. The red and green framed portions are both locally enlarged images. This display method allows for a more intuitive comparison of the clarity of pCLE images.

3.5. Comparisons to Other Deep Learning Deblurring Methods

To better evaluate the proposed algorithm's performance, we recovered images on our dataset and compared them with other algorithms, including Deblur-GAN [26], DGUNet [31], and MIMO-Unet [32]. An intuitive comparison is shown in Figure 7. Figure 7f are the ground truth images from the test set. Then we first added various fuzziness to obtain Figure 7a, and we used Figure 7a as input for each algorithm to obtain each of the remaining images in Figure 7, respectively. Compared with the previous image processing methods, it can be seen that this method can obviously recover the spatial details of W2S images better. The PSNR and SSIM values calculated for each image and the ground truth image are listed below the image, respectively. The PSNR and SSIM of the proposed algorithm exceeded that of Deblur-GAN, DGUNet, and MIMO-Unet. The tangible performance gains, quantified through metrics such as PSNR and SSIM, underscore the efficacy of our approach in significantly enhancing image quality [33]. Notably, the visual results further emphasize the ability of our method to produce images that are not only sharper but also more faithful to the original scenes [34], crucial qualities in applications such as medical imaging, surveillance, and other domains where precision and accuracy are paramount.

We tested the deblurring effect of various methods on a dataset of 120 images as a whole. Table 1 shows the average quantification index of 120 images in terms of PSNR and SSIM. Obviously, for the three indicators of MSE, PSNR, and SSIM, the proposed method is significantly improved compared with the previous method, and compared with MIMO-Unet, the improvements are 22.427, 1.113 (dB), and 0.023, respectively.

Table 1. Comparison of quantitative indicators.

Algorithm	MSE	PSNR/dB	SSIM
Previous work	93.845	27.979	0.817
Deblur-GAN	860.423	28.359	0.828
DGUNet	83.300	28.463	0.828
MIMO-Unet	82.140	28.530	0.832
Ours	59.713	29.643	0.855

The robustness of our method is particularly evident in its ability to adapt to varying data characteristics, demonstrating consistent performance across synthetic datasets that simulate controlled scenarios and real datasets that mirror the complexity of authentic environment. This versatility underscores the practical applicability of our deblurring method across a range of real-world situations [35]. In addition, we test the deblurring effect of this method on fuzzy images in real life scenes. This part of the dataset has 1111 images. The 1111 images were tested under three conditions, and the average PSNR/SSIM values we calculated are shown in Table 2 below. Our proposed method achieves a performance

gain of 0.72 dB/0.319 dB on the dataset. It can be seen that this method also has good deblurring effect for common datasets and can achieve high-resolution recovery.

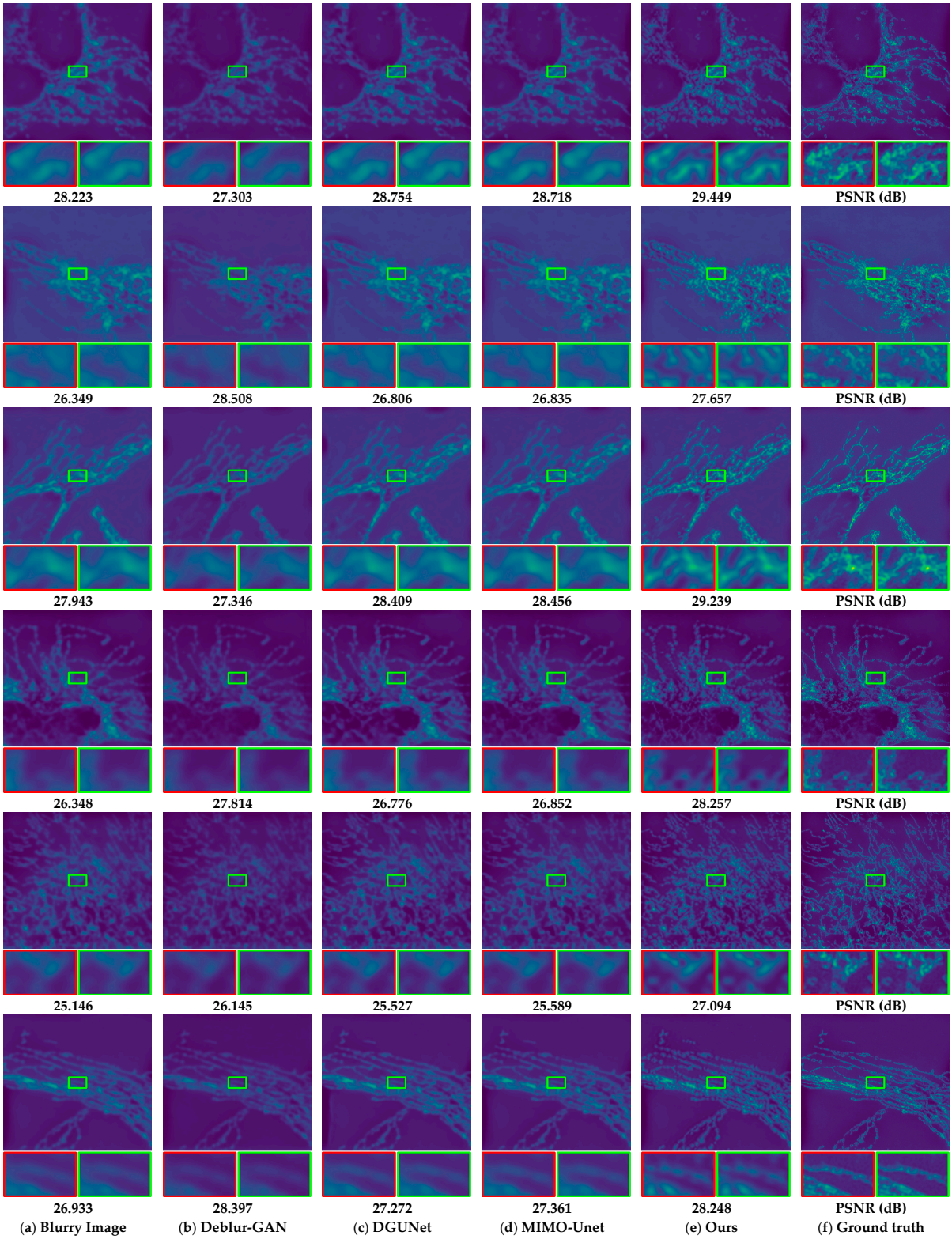


Figure 7. The pCLE deblurring results obtained using Deblur-GAN, DGUNet, MIMO-Unet, and the proposed method. The red and green framed portions show magnified sections, allowing for a more intuitive comparison of the clarity of different images.

Table 2. The average PSNR/SSIM value of the GoPro deblurring image.

Dataset	CAM	AM Filter	PSNR/SSIM
GoPro	√	×	28.393/0.859
GoPro	×	√	27.992/0.851
GoPro	√	√	28.712/0.868

3.6. Ablation Studies

Here we present the contribution of ablation experiments to analyze the important component of our model. In this method, we give priority to CAM architecture, combined with AM filter, which is easy to confuse the concept and meaning of the two. CAM is a cross-channel attention module introduced between every two stages that is capable of aggregating local and nonlocal pixel interactions and is efficient enough to process high-resolution images. The global context is implicitly emulated by applying self-attention across channels, resulting in linear complexity rather than quadratic complexity. The AM filter, functioning as a supervised attention module, furnishes a valuable truth-monitoring signal for progressive image recovery at each stage. Additionally, leveraging locally supervised prediction, it can generate attention maps to suppress less informative features in the current stage, allowing only pertinent features to propagate to the subsequent stage. In fact, they have different functions and effects, but they are all attention modules, which process different aspects of image information features in turn, so as to achieve a better deblurring optimization effect. We have tested the optimization effects of the two parts, respectively, which are divided into the following three situations: there is CAM, there is no AM filter; there is no CAM, there is AM filter; both CAM and AM filter are available. The results are shown in Table 3 below. The dataset is W2S, with a total of 120 images, and the quantified PSNR/SSIM scores are the average value.

Table 3. Ablation study on individual components of the proposed method.

Dataset	CAM	AM Filter	PSNR/SSIM
W2S	√	×	28.579/0.833
W2S	×	√	28.363/0.826
W2S	√	√	29.643/0.855

We demonstrate the effectiveness of the proposed CAM and AM filter mechanism by removing them from our final model. Table 3 shows a substantial drop in PSNR from 29.643 dB to 28.363 dB when CAM is removed, and from 29.643 dB to 28.579 dB when we take out AM filter.

4. Discussion

The presented work introduces a novel approach to image deblurring by adding cross-channel attention module (CAM) into the multistage neural networks' architecture. The application of this advanced method to the deblurring of high-resolution pCLE images demonstrates notable improvements in both quantitative and qualitative aspects. In this discussion, we delve into the significance of the introduced CAM, the effectiveness of the proposed method, and its potential implications for the field of image deblurring. The cross-channel attention module (CAM) serves as a pivotal element in our proposed deblurring method. By enhancing the ability to capture multiscale deep textures and preserving structural informat ion, the CAM addresses the intricate challenges associated with deblurring high-resolution pCLE images. The utilization of this module in multiple stages of the proposed approach ensures a comprehensive refinement process, allowing for the aggregation of finer details and the production of high-quality, visually faithful images. The iterative refinement process employed in our approach stands out as a key factor contributing to the enhanced restoration quality. At the current stage, our proposed method is applied to deblurring in image restoration, while its effectiveness in other aspects

of image restoration, such as denoising and deraining, still requires improvement. We will further improve our method and strive to achieve better optimization results. At present, it is mainly divided into the following five directions: Explore the possibility of developing an adaptive mechanism for tuning the parameters of the cross-channel attention module (CAM) based on the characteristics of specific images. Investigate methods to incorporate temporal information into the deblurring process, especially for applications involving dynamic scenes. Explore the integration of information from multiple modalities, such as additional imaging modalities or ancillary data sources. Train the deblurring model on larger and more diverse datasets to further improve its generalization capabilities. A more extensive training dataset could enable the model to better adapt to a broader range of image characteristics, enhancing its performance across varied real-world scenarios [36,37]. Extend the evaluation of the proposed deblurring method to diverse applications beyond medical imaging, such as surveillance, satellite imagery, or artistic content. Benchmarking the algorithm in various contexts will provide insights into its versatility and potential extensions.

5. Conclusions

In this work, we propose the concept of cross-channel attention module (CAM). We introduce an advanced method for deblurring high-resolution pCLE (probe-based confocal laser endomicroscopy) images by enhancing the cross-channel attention module (CAM). Our proposed approach comprises multiple stages, each dedicated to enhancing the restored image through the utilization of the CAM. In the initial stage, conventional upscaling techniques are applied to the low-resolution input image. Subsequent stages leverage the CAM to aggregate multiscale deep textures, capturing intricate details and preserving structural information. The multistage neural networks contribute to improving the quality of restoration, resulting in high-resolution images with improved visual fidelity. Due to its multistage nature, challenging image restoration tasks can be decomposed into subtasks, enabling the progressive recovery of degraded images. The experimental results demonstrate the superiority of our proposed algorithm over the comparison algorithm in both quantitative and qualitative analyses of synthetic and real datasets. The restored images exhibit a higher level of clarity, retaining more detailed information. In summary, our evaluation showcases the robustness and efficacy of the proposed deblurring method across synthetic and real datasets. The tangible performance gain, as evidenced by the quantitative metrics and visual results, positions our approach as a promising solution for addressing the challenges associated with image deblurring in diverse and authentic environments.

Author Contributions: Conceptualization, H.Z. and Y.Q.; methodology, H.Z., Y.Q. and K.Y.; software, T.Z. and Y.L.; validation, Z.C.; data curation, Z.Z.; writing—original draft preparation, Y.Q.; writing—review and editing, H.Z.; project administration, H.Z.; funding acquisition, H.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This work was 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. Shao, J.; Zhang, J.; Liang, R.; Barnard, K. Fiber bundle imaging resolution enhancement using deep learning. *Opt. Express* **2019**, *27*, 15880–15890. [CrossRef] [PubMed]
2. Hughes, M.; Yang, G.Z. Line-scanning fiber bundle endomicroscopy with a virtual detector slit. *Biomed. Opt. Express* **2016**, *7*, 2257–2268. [CrossRef] [PubMed]

3. Thrapp, A.D.; Hughes, M.R. Automatic Motion Compensation for Structured Illumination Endomicroscopy Using a Flexible Fiber Bundle. *J. Biomed. Opt.* **2020**, *25*, 026501. [CrossRef] [PubMed]
4. Zhang, K.; Chen, M.; Zhu, D.; Liu, K.; Zhao, H.; Liao, J. Multi-Scale Cyclic Image Deblurring Based on PVC-Resnet. *Photonics* **2023**, *10*, 862. [CrossRef]
5. Perperidis, A.; Dhaliwal, K.; McLaughlin, S. Image computing for fibre-bundle endomicroscopy: A review. *Med. Image Anal.* **2020**, *62*, 101620. [CrossRef]
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. Weigert, M.; Schmidt, U.; Boothe, T. Content-aware image restoration: Pushing the limits of fluorescence microscopy. *Nat. Methods* **2018**, *15*, 1090–1097. [CrossRef]
8. Ma, H.; Zhang, W.; Ning, X.; Liu, H.; Zhang, P.; Zhang, J. Turbulence Aberration Restoration Based on Light Intensity Image Using GoogLeNet. *Photonics* **2023**, *10*, 265. [CrossRef]
9. Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative Adversarial Networks. *Commun. ACM* **2020**, *63*, 139–144. [CrossRef]
10. Huang, Z.; Wang, X.; Huang, L.; Huang, C.; Wei, Y.; Liu, W. CCNet: Criss-Cross Attention for Semantic Segmentation. In Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision (ICCV), Seoul, Republic of Korea, 27 October–2 November 2019.
11. Tao, X.; Zhou, H.; Chen, Y. Image Restoration Based on End-to-End Unrolled Network. *Photonics* **2021**, *8*, 376. [CrossRef]
12. Guo, Y.; Wu, X.; Qing, C.; Su, C.; Yang, Q.; Wang, Z. Blind Restoration of Images Distorted by Atmospheric Turbulence Based on Deep Transfer Learning. *Photonics* **2022**, *9*, 582. [CrossRef]
13. Pan, J.S.; Dong, J.X.; Liu, Y.; Zhang, J.W.; Ren, J.M.; Tang, J.H.; Tai, Y.W.; Yang, M.H. Physics-Based Generative Adversarial Models for Image Restoration and Beyond. *IEEE Trans. Pattern Anal. Mach. Intell.* **2021**, *43*, 2449–2462. [CrossRef]
14. Zhang, K.; Zuo, W.; Chen, Y.; Meng, D.; Zhang, L. Beyond a Gaussian Denoiser: Residual Learning of Deep CNN for Image Denoising. *IEEE Trans. Image Process.* **2017**, *26*, 3142–3155. [CrossRef] [PubMed]
15. Mao, X.J.; Shen, C.; Yang, Y.B. Image restoration using very deep convolutional encoder-decoder networks with symmetric skip connections. In Proceedings of the 30th International Conference on Neural Information Processing Systems (NIPS'16), Barcelona, Spain, 5–10 December 2016.
16. Liu, M.; Su, X.; Yao, X.; Hao, W.; Zhu, W. Lensless Image Restoration Based on Multi-Stage Deep Neural Networks and Pix2pix Architecture. *Photonics* **2023**, *10*, 1274. [CrossRef]
17. Cheng, J.; Zhu, W.; Li, J.; Xu, G.; Chen, X.; Yao, C. Restoration of Atmospheric Turbulence-Degraded Short-Exposure Image Based on Convolution Neural Network. *Photonics* **2023**, *10*, 666. [CrossRef]
18. Asim, M.; Shamshad, F.; Ahmed, A. Blind Image Deconvolution Using Deep Generative Priors. *IEEE Trans. Comput. Imaging* **2020**, *6*, 1493–1506. [CrossRef]
19. Xue, J.; Liang, J.; He, J.; Zhang, Y.; Hu, Y. MMPDNet: Multi-Stage & Multi-Attention Progressive Image Denoising. In Proceedings of the 2021 20th International Conference on Ubiquitous Computing and Communications (IUCC/CIT/DSCI/SmartCNS), London, UK, 20–22 December 2021.
20. Zamir, S.W.; Arora, A.; Khan, S.; Hayat, M.; Khan, F.S.; Yang, M.H. Restormer: Efficient Transformer for High-Resolution Image Restoration. In Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), New Orleans, LA, USA, 18–24 June 2022.
21. Zamir, S.W.; Arora, A.; Khan, S.; Hayat, M.; Khan, F.S.; Yang, M.H.; Shao, L. Multi-stage progressive image restoration. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 20–25 June 2021; pp. 14821–14831.
22. 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]
23. Anwar, S.; Khan, S.; Barnes, N. A deep journey into super-resolution: A survey. *ACM Comput. Surv.* **2019**, *53*, 1–34. [CrossRef]
24. Aubreville, M.; Stoeve, M.; Oetter, N.; Goncalves, M.; Knipfer, C.; Neumann, H.; Bohr, C.; Stelzle, F.; Maier, A. Deep learning-based detection of motion artifacts in probe-based confocal laser endomicroscopy images. *Int. J. CARS* **2019**, *14*, 31–42. [CrossRef]
25. Chen, L.; Bentley, P.; Mori, K.; Misawa, K.; Fujiwara, M.; Rueckert, D. Self-supervised learning for medical image analysis using image context restoration. *Med. Image Anal.* **2019**, *58*, 101539. [CrossRef]
26. Kupyn, O.; Budzan, V.; Mykhailych, M.; Mishkin, D.; Matas, J. DeblurGAN: Blind Motion Deblurring Using Conditional Adversarial Networks. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, 18–23 June 2018.
27. Zhang, Y.; Zhu, Y.; Nichols, E.; Wang, Q.; Zhang, S.; Smith, C.; Howard, S. A Poisson-Gaussian Denoising Dataset with Real Fluorescence Microscopy Images. In Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 15–20 June 2019.
28. 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 Computer Vision—ECCV 2020 Workshops, Glasgow, UK, 23–28 August 2020.

29. 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]
30. Chan, T.; Esedoglu, S.; Park, F.; Yip, A. Total variation image restoration: Overview and recent developments. In *Handbook of Mathematical Models in Computer Vision*; Springer: Berlin/Heidelberg, Germany, 2006; pp. 17–31.
31. Mou, C.; Wang, Q.; Zhang, J. Deep Generalized Unfolding Networks for Image Restoration. In Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), New Orleans, LA, USA, 18–24 June 2022.
32. Cho, S.-J.; Ji, S.-W.; Hong, J.-P.; Jung, S.-W.; Ko, S.-J. Rethinking Coarse-to-Fine Approach in Single Image Deblurring 2021. In Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision (ICCV), Montreal, QC, Canada, 10–17 October 2021.
33. Liehr, S.; Borchardt, C.; Münzenberger, S. Long-Distance Fiber Optic Vibration Sensing Using Convolutional Neural Networks as Real-Time Denoisers. *Opt. Express* **2020**, *28*, 39311. [CrossRef]
34. Zhang, K.; Zuo, W.; Zhang, L. FFDNet: Toward a Fast and Flexible Solution for CNN-Based Image Denoising. *IEEE Trans. Image Process.* **2018**, *27*, 4608–4622. [CrossRef] [PubMed]
35. Shan, Q.; Jia, J.; Agarwala, A. High-quality motion deblurring from a single image. *ACM Trans. Graph.* **2008**, *27*, 1–10.
36. Yarotsky, D. Error Bounds for Approximations with Deep ReLU Networks. *Neural Netw.* **2017**, *94*, 103–114. [CrossRef] [PubMed]
37. Adams, J.K.; Boominathan, V.; Avants, B.W.; Vercosa, D.G.; Ye, F.; Baraniuk, R.G.; Robinson, J.T.; Veeraraghavan, A. Single-frame 3D fluorescence microscopy with ultraminiature lensless FlatScope. *Sci. Adv.* **2017**, *3*, e1701548. [CrossRef]