Complex & Intelligent Systems (2025) 11:55
https://doi.org/10.1007/s40747-024-01633-7

ORIGINAL ARTICLE

UNet-Att: a self-supervised denoising and recovery model for two-photon microscopic image

Yuer Lu[1,2] · Yongfa Ying[3] · Chen Lin[3] · Yan Wang[4] · Jun Jin[3] · Xiaoming Jiang[1] · Jianwei Shuai[1,2] · Xiang Li[3] · Jinjin Zhong[1,2]

Received: 24 January 2024 / Accepted: 1 September 2024

Abstract

Two-photon microscopy is indispensable in cell and molecular biology for its high-resolution visualization of cellular and molecular dynamics. However, the inevitable low signal-to-noise conditions significantly degrade image quality, obscuring essential details and complicating morphological analysis. While existing denoising methods such as CNNs, Noise2Noise, and DeepCAD serve broad applications in imaging, they still have limitations in preserving texture structures and fine details in two-photon microscopic images affected by complex noise, particularly in sophisticated structures like neuronal synapses. To improve two-photon microscopy image denoising effectiveness, by experimenting on real two-photon microscopy images, we propose a novel deep learning framework, the UNet-Att model, which integrates a specifically tailored UNet++ architecture with attention mechanisms. Specifically, this approach consists of a sophisticated downsampling module for extracting hierarchical features at varied scales, and an innovative attention module that strategically emphasizes salient features during the integration process. The architecture is completed by an ingenious upsampling pathway that reconstructs the image with high fidelity, ensuring the retention of textural integrity. Additionally, the model supports end-to-end training, optimizing its denoising efficacy. The UNet-Att model proves to surpass mainstream algorithms in the dual objectives of denoising and preserving the textural intricacies of original images, which is evidenced by an increase of 9.42 dB in the high Peak Signal-to-Noise Ratio (PSNR) coupled with an improvement of 0.1131 in the Structural Similarity Index Measurement (SSIM). The ablation experiments reveal the effectiveness of each module. The associated Python packages and datasets of UNet-Att are freely available at https://github.com/ZjjDh/UNet-Att.

Keywords Two-photon microscopy image · Image denoising · Deep learning · Attention · Unet++

Introduction

Two-photon microscopy [1–3] is a powerful imaging technique that combines the depth-penetrating prowess of laser scanning confocal microscopy [4] with the precision of two-photon excitation. It has emerged as a potent tool for elucidating cellular structures and functions in exquisite detail, enabling deep tissue visualization and providing superior imaging resolution compared to conventional single-photon methods [5]. This imaging technique is particularly valuable for studying cellular activities and tracking changes within individual cells, thereby advancing the comprehension of cellular physiology. Additionally, two-photon microscopy offers a reduced risk of phototoxicity and cellular damage, making it suitable for prolonged observation of living cells, including neurons [6]. When combined with calcium-dependent fluorescent indicators, it allows for capturing neuronal

Yuer Lu and Yongfa Ying contributed equally to this work.

✉ Xiang Li
xianglibp@xmu.edu.cn

✉ Jinjin Zhong
zjj89@wiucas.ac.cn

1 Wenzhou Key Laboratory of Biophysics, Wenzhou Institute, University of Chinese Academy of Sciences, Wenzhou, Zhejiang 325001, China

2 Oujiang Laboratory (Zhejiang Lab for Regenerative Medicine, Vision and Brain Health), Wenzhou, Zhejiang 325001, China

3 Department of Physics, and Fujian Provincial Key Laboratory for Soft Functional Materials Research, Xiamen University, Xiamen 361005, China

4 McCormick School of Engineering, Northwestern University, Evanston, IL 60208, USA

Published online: 26 November 2024

dynamics and producing video recordings of neuronal activity. In live imaging, there is often a trade-off between light exposure and image quality to avoid photobleaching and phototoxicity. Tracking very rapid processes usually necessitates short exposure times and low-dose fluorescence dyes, resulting in low signal-to-noise (SNR) images. Therefore, addressing the challenges of noise reduction and signal recovery in low SNR images is crucial for improving image quality in two-photon microscopy live image processing.

Various methods have been developed for noise reduction and signal recovery in fluorescence microscopic images. Traditional approaches include filtering-based methods such as WNNM [7], TNRD [8], EPLL [9], NLM [10], and BM3D [11], as well as mathematical optimization methods [12] such as the KSVD [13] method. Traditional denoising methods have very limited applicability due to their unsatisfactory performance under complex noise conditions [11, 14]. In recent years, deep learning-based algorithms [15–19] have gained prominence due to the increased computational power of graphics processing units and their remarkable denoising performance [20–24]. Deep learning-based approaches do not require explicit analytical modeling of prior knowledge or noise models. Instead, they rely on specific patterns learned from training datasets obtained from experiments or simulations to achieve end-to-end image transformation [25]. The learned patterns depend heavily on the image features present in the training dataset, enabling the neural network model to deliver optimal noise reduction results for data similar to the training dataset.

Deep learning methods first applied to image denoising in the 1980s used neural networks with known translationally invariant blur functions and additive noise to produce denoised images [26]. Jain et al. (2008) [27] proposed a four-layer neural network that achieved denoising effects comparable to traditional methods. Xie et al. (2012) [28] introduced Stacked Sparse Denoising Autoencoders (SSDA) for unsupervised image denoising, combining sparse coding with deep neural networks. Burger et al. (2012) [29] proposed a Multilayer Perceptron (MLP) model for image denoising, effectively removing various types of noise.

The advent of Convolutional Neural Networks (CNNs) marked a new era for deep learning-based image denoising [30]. CNNs achieved speeds and accuracies in image processing tasks that rivaled human visual recognition, significantly surpassing the limitations of previous traditional algorithms. CNNs, through their properties of translational invariance, local connections, and weight sharing, demonstrated considerable advantages in 2D image processing. Increasing the width and depth of networks allowed for the extraction of higher-dimensional and deeper feature maps, enhancing denoising performance. However, increasing the number of neural network layers introduced training issues such as gradient explosion and vanishing gradients. Mao et al. [31] proposed a fully convolutional encoder-decoder network for image denoising and restoration. This network, composed of symmetric multi-layer convolutional and deconvolutional operators, learns the end-to-end mapping from noisy images to original images. Since then, the depth of neural networks for image denoising has continued to grow.

In the field of fluorescence microscopy image denoising, CNNs have also outperformed traditional methods and have gained great development, giving rise to numerous outstanding models. A substantial portion of these models is based on the U-Net architecture [32], which integrates deep and shallow features through concatenation, effectively handling complex and data-sparse biomedical images. U-Net consists of an encoder, a decoder, and skip connections between them. In the encoding phase, U-Net reduces the resolution through downsampling to capture multi-scale image information. In this process, image information gradually transforms from point and line information in the low-level information to contour and more abstract semantic information in the high-level information. The entire network completes the process of feature extraction and combination from fine to coarse, enabling U-Net to obtain more comprehensive information. In the decoding layer operation, the original image resolution needs to be restored. During the process of converting the image from low resolution to high resolution, the image will be distorted, and the upsampling process will also lose sensitivity to detail information. The addition of skip connections transmits the detail information of the shallow feature map to the deep feature map, enabling the network to better recognize small targets and edge information. In addition, during the decoding layer stage, feature extraction is performed at different scales, and the features extracted later will cover the features extracted earlier to a certain extent, causing loss of detail information and feature information dilution. Feature fusion through skip connections can replenish the diluted information, allowing the network to retain more contextual information. U-Net has become a widely used neural network in biomedical imaging, achieving excellent results in many image processing tasks including image segmentation and denoising [32, 33].

Traditional deep learning-based denoising approaches were developed as supervised methods [34] which involve a set of corresponding noisy input images and clean target images. The great challenge of obtaining paired noisy and noise-free image datasets in fluorescence microscopy has given rise to self-supervised learning [35, 36], which only requires noisy images for training. This strategy leverages the characteristics of fluorescence microscopy images, where long-term noise-free exposure is equivalent to a series of short, independent noisy exposures. Despite the

advancements in deep learning-based denoising, existing methods for the image denoising tasks with sophisticated hierarchical structures still have limitations in recovering structural details at different scales at the same time. The limitations include oversmoothing, loss of structural details and image contrast information, when relatively complex hierarchical structures are involved in the images. CNNs may suffer from overfitting and lack of generalization when trained on limited or biased data, and exhibit a spectral bias, which predisposes them to favor low-frequency features while neglecting high-frequency details [37]. This bias often results in overly smoothed denoising outcomes, which can compromise the integrity of the original image features, especially for sophisticated hierarchical structures.

In this work, we propose a self-supervised method, UNet-Att, which integrates a specifically designed UNet++ network with advanced attention mechanisms for denoising synchronized two-photon calcium imaging datasets with low signal-to-noise ratios [38]. This method effectively removes image noise while preserving image texture features, outperforming existing mainstream algorithms with an increase of 9.42 dB in PSNR and an improvement of 0.1131 in SSIM. To improve accuracy in images with complex hierarchical structures, we devised a specific advanced attention mechanism into our specially designed network based on the UNet++ framework. This approach aligns with the need to improve denoising effectiveness while preserving structural details. The main contribution of this research can be summarized as follows:

1) To enhance the image denoising effectiveness for complex hierarchical structures, we combine the architectural strength of UNet++ with attention mechanism.
2) To ensure the robust feature representation, we utilize a down-sampling module to extract image features at multiple scales.
3) To improve the preservation of important features for the hierarchical structures, we employ attention modules to emphasize the weights of structural information in the image.
4) To ensure the comprehensive reconstruction, we implement a multi-dimensional upsampling module to restore image information and merge the outputs from various network dimensions.
5) To obviate the difficulties of obtaining the counterpart ground-truth images, we successfully apply self-supervised training on three-dimensional low SNR two-photon microscopy images.

The remainder of this paper proceeds as follows: Sect. 2 provides a literature review of the related denoising methods. Section 3 elaborates on the proposed self-supervised deep learning denoising method, UNet-Att, including the noise model, the architecture of the deep neural network, and the denoising mechanism. Section 4 details the application of UNet-Att in denoising the two-photon microscopic images, and benchmark the denoising effectiveness for the live neuronal imaging with a systematic comparative study. Finally, Sect. 5 offers conclusions and outlines prospective avenues for future investigations.

Related work

Advanced algorithms based on different networks have been developed to further enhance performance in different aspects and for various application requirements, including reducing computation time, improving algorithm generalization, and enhancing accuracy for specific tasks. There are often trade-offs between accuracy, generalization, and computation time. For applications that demand high accuracy in recovering structural details and maintaining the integrity of original image features during denoising, specifically tailored algorithms are being prioritized and extensively explored for corresponding applications. In this section, we introduce previous related work in image denoising methods based on deep learning, which is also summarized in Table 1 with the proposed method in this work.

Supervised image denoising based on CNNs

Supervised image denoising methods using CNNs have been extensively explored in the realm of biofluorescent microscopic imaging. CNNs exhibit significant advantages in the realm of two-dimensional image processing, primarily due to their properties of translational invariance, local connectivity, and weight sharing [30]. These characteristics contribute to the simplicity of computations, reduced complexity, and the ability to leverage GPU acceleration for faster processing. Furthermore, increasing the width and depth of CNNs enables the extraction of higher-dimensional and deeper features, thus enhancing denoising performance. However, the expansion of neural network layers can lead to training challenges such as gradient explosion and gradient vanishing.

To address these issues, Ioffe et al. [39] introduced Batch Normalization (BN), which stabilizes the training process by normalizing the inputs of each layer. Additionally, He et al. [40] proposed the ResNet model, incorporating residual connections to effectively mitigate these training difficulties, thereby easing the training of deeper networks. These innovations marked a significant maturation of CNN architectures, paving the way for more sophisticated models. Mao et al. [31] developed a fully convolutional encoder-decoder

Table 1 Related image denoising methods

Method	Approach	Focus	Key innovations	Contributions
DnCNN (Zhang et al.) [41]	Supervised CNN	Regular 2D image denoising	Residual learning, batch normalization, and blind denoising	PSNR improvement over traditional denoising methods
Mu-net (Lee et al.) [43]	Supervised3D U-Net	Two-photon microscopic image denoising	End-to-end learning	Effective denoising with limited ground-truth images
CARE (Weigert et al.) [42]	Supervised CNN	Fluorescence microscopic image restoration	Content-aware neural networks	High-quality denoising with clean references
Noise2Noise (Lehtinen et al.) [38]	Self-supervised U-Net	General 2D image restoration	Training without clean data	Foundation for self-supervised methods
Noise2Fast (Lequyer et al.) [51]	Self-supervised CNN	Real-time denoising in fluorescence microscopy	A novel chequerboard downsampling technique	Higher-speed denoising
DeepCAD (Li et al.) [36]	Self-supervised 3D U-Net	Spatiotemporal enhancement of calcium imaging	Temporal redundancy in 3D	Improved restoration accuracy for neuron extraction and spike inference
DeepCAD-RT (Li et al.) [52]	DeepCAD	Real-time denoising in fluorescence microscopy	Model simplification and data augmentation	Significant reduction in denoising time
UniFMIR (Ma et al.) [37]	Transfer learning	Fluorescence microscopic image restoration	Multihead and multitail network architecture, and swin transformer-based feature enhancement module	Improved generalization and increased versatility
SRDTrans (Li et al.) [53]	Self-supervised, lightweight spatiotemporal transformer	Fluorescence image SNR enhancement	Spatial redundancy sampling strategy and lightweight spatiotemporal transformer architecture	High-frequency information restoration
UNet-Att (This work)	Self-supervised, UNet++ with attention	Denoising of complex hierarchical structures	Specifically designed hierarchical feature preservation with attention mechanisms	Improved denoising accuracy for low SNR images with complex hierarchical structures

network for image denoising and restoration. This network consists of symmetrical convolutional and deconvolutional operators, learning an end-to-end mapping from noisy images to clean images. This architecture laid the foundation for deeper neural networks in image denoising. Building on this progress, Zhang et al. [41] introduced the DnCNN model, one of the most renowned CNN-based algorithms for image denoising. The network employs a series of convolutional layers to learn the residual noise in images. By focusing on the residual noise, DnCNN achieves superior denoising performance, surpassing all traditional denoising methods of its time.

While DnCNN and similar models have achieved remarkable performance, they often struggle with preserving intricate details and textures in highly noisy images. Additionally, the increased network depth and complexity can result in higher computational demands and longer training times. These supervised methods are inherently limited by their reliance on ground-truth images, which are often difficult to obtain in practice.

U-net-derived supervised denoising for fluorescence microscopic images

In the domain of fluorescence microscopy image denoising, CNNs have also outperformed traditional image denoising methods. Ronneberger et al. [32] introduced the renowned U-Net architecture. U-Net employs a concatenation technique to merge deep and shallow features through multiscale information extraction and skip connections. This design allows U-Net to effectively handle complex biomedical images, even with limited data, making it a staple in the biomedical imaging field. Its versatility and superior performance in various tasks have cemented U-Net as a classic neural network model in this domain.

Building on the success of U-Net, Weigert et al. [42] proposed the Content-aware Image Restoration (CARE)

method. CARE is a task-specific network framework based on U-Net, offering a user-friendly interface for researchers to perform fluorescence microscopy image denoising. This approach exemplifies the practical application of U-Net in enhancing the usability and effectiveness of image restoration tools. Lee et al. [43] introduced the Mu-net method, which involves the sequential use of multiple U-Net networks combined with a Generative Adversarial Network (GAN) [44]. This hybrid approach achieved remarkable results in denoising three-dimensional two-photon microscopy images, demonstrating the potential of combining multiple neural network architectures for improved performance.

All these methods rely on supervised training strategies to achieve effective image denoising. Supervised training typically yields excellent results by utilizing paired datasets of noisy input images and clean target images [34]. In the field of fluorescence microscopy, obtaining paired datasets of noisy and noise-free images is challenging. Additionally, the performance of supervised deep learning models is largely dependent on the characteristics of the training dataset. While these models excel with data similar to the training set, their performance often degrades when applied to significantly different data. This limitation underscores the need for methods that can generalize better across diverse datasets.

Self-supervised image denoising approaches

To obviate the need for clean ground-truth image, self-supervised image denoising methods have emerged as a powerful alternative [45–47]. In the domain of fluorescence microscopy image denoising, several innovative methods have employed self-supervised learning strategies to train their models effectively. Lehtinen et al. [36, 38] introduced the pioneering Noise2Noise method, which describes a self-supervised training approach for denoising. Unlike supervised methods, Noise2Noise does not require clean images as targets for neural network learning. Instead, it only requires noisy images for training. This strategy is particularly advantageous in the context of fluorescence microscopy, where a long noise-free exposure can be approximated by a series of short, independent noisy exposures. Under this framework, even though the neural network learns from noisy images, it can still extract the clean signal from a series of temporally stacked noisy images.

Building on this concept, Li et al. [36] proposed the DeepCAD method. DeepCAD utilizes a three-dimensional U-Net architecture, significantly improving the signal-to-noise ratio of two-photon microscopy images by more than tenfold. Additionally, this method enhances the accuracy of subsequent image processing tasks, such as neuron extraction and spike inference. In the context of calcium imaging, DeepCAD explores the temporal redundancy inherent in video-rate imaging. It posits that any two consecutive frames can be considered independent samples of the same underlying discharge pattern. Consequently, image pairs composed of consecutive frames can be utilized for training the denoising model. However, the necessity for a specialized imaging system to generate high and low SNR pairs may not be practical for all research settings, limiting the method's general applicability.

Attention mechanisms in deep learning for image denoising

Attention mechanisms have been successfully integrated into deep learning models to enhance performance in various applications, including image denoising. These mechanisms enable models to focus on significant parts of the data and thereby enhance their performance in critical feature extraction and preservation. Attention mechanisms are typically categorized into two types: hard attention and soft attention. Hard attention, as described by Xu et al. [48], involves the model focusing on a fixed position within the input sequence. In this type of attention, the model selects a specific location in the input sequence as the focal point, extracting necessary information from this position while ignoring other locations. However, the non-differentiable nature of hard attention makes it challenging to optimize during training, thereby limiting its applicability. In contrast, soft attention allows the model to consider all input positions, assigning different weights to each position. These weights indicate the importance of the information at each location, enabling the model to focus on various parts of the input data with varying intensities. The output of the model in soft attention is a weighted sum of all input positions. Soft attention is more commonly used due to its differentiable nature, facilitating easier optimization during training.

Several studies have successfully employed soft attention mechanisms to enhance image processing tasks. Xiao et al. [35] introduced a spatial transformer module that transforms spatial domain information within images to extract key features. SENet [49] focuses on the weight relationships between channels, enhancing model performance by increasing the weights of important channels and reducing those of less significant ones. Oktay et al. [50] applied attention mechanisms to skip connections, boosting the importance of structural features within the images.

Other advanced algorithms based on the above work have been developed to further enhance performance in different aspects and for various application requirements, including reducing the computation time and improving

the algorithm generalization. Examples include Noise2Fast [51], DeepCAD-RT [52] for faster denoising with reduced requirement of computational resources and UniFMIR [37] for improving the generalization ability. There are often trade-offs between the accuracy, generalization, and computation time. For applications that demand high accuracy in recovering structural details and maintaining the integrity of original image features during denoising, specifically tailored algorithms are being prioritized and extensively explored for corresponding applications [53]. For synchronized two-photon calcium imaging datasets with low SNR and complex hierarchical structures, existing methods still face limitations in preserving the structural details in denoising.

UNet++ is a notable enhancement of the original U-Net architecture, involving the integration of multiple U-Net models to achieve comprehensive feature fusion at each layer's output. This work aims to enhance the denoising effectiveness of the calcium imaging with low SNR and complex hierarchical structures. By integrating specially designed UNet++ architecture with advanced attention mechanisms, our proposed method, UNet-Att, achieves a better performance than mainstream denoising algorithms.

Method

This study is dedicated to proposing a deep learning-based method for enhancing the effectiveness of noise reduction and signal recovery in low SNR images of sophisticated hierarchical structures obtained from two-photon microscopy live imaging. To achieve this, the noise model was firstly described. Based on the insight of the noise characteristics, the research develops a self-supervised method taking advantage of the powerful capability of hierarchical feature representation of UNet++ network and incorporating advanced attention mechanisms which augment the denoising effectiveness. The proposed network was named as UNet-Att.

Noise model

Two-photon microscopy images frequently encounter significant noise issues. These issues predominantly arise from two sources: quantum shot noise and detector-related noise, both of which have distinct origins and characteristics. Quantum shot noise is an inherent property of the photon detection process, following a Poisson distribution mathematically. This type of noise is signal-dependent, resulting in higher noise levels in brighter regions due to increased photon counts [54]. Conversely, detector noise, which takes a Gaussian form, impacts each pixel independently and consistently across the image, irrespective of the pixel's actual intensity [55].

The experimentally detected signal at the pixel i for the two-photon microscopy imaging process can be represented as [56]:

$$X_i = a\varphi\ (S_i) + \epsilon_i \tag{1}$$

where X_i is the measured signal, S_i is the underlying true signal, and $\phi\ (S_i)$ denotes the signal under the influence of shot noise. a denotes the conversion factor, and ϵ_i denotes the detector noise. The shot noise-affected signal intensity $\phi\ (S_i)$ adheres to a Poisson distribution centered around S_i and randomly fluctuates by a standard deviation of $\sqrt{S_i}$. The detector noise ϵ_i, is generally characterized by a Gaussian distribution with a standard deviation that remains constant and is independent of the underlying signal. The Gaussian distribution is detector-dependent, and is centered around a baseline offset. The offset is pre-configured by most camera systems to preclude the occurrence of negative values in the Gaussian noise, which are inherently non-physical, and this offset is consequently superimposed onto the measured signal. Both the quantum shot noise and the detector noise occur independently for each pixel without influencing its neighbors.

It is evident that in the presence of shot noise and detector noise, the signal acquired during each measurement is subject to fluctuate around the true signal. Over multiple acquisitions, the mean of these measurements is expected to converge towards the true signal. This characteristic can be harnessed using a self-supervised deep learning methodology wherein a collection of image stack data is bifurcated into two distinct sets, predicated on individual frames. Each corresponding frame in these sets encapsulates identical underlying biological information, albeit affected by disparate noise intensities. By feeding these image stacks into a neural network as input and target data for learning, the network is trained to discern the average content within the stacks. This process enables the neural network to effectively perform image denoising by learning to filter out the noise and recover the signal of interest.

Self-supervised method

By fully accounting for the distinct noise attributes inherent in two-photon microscopy images, the proposed UNet-Att was a sophisticated deep learning algorithm for image denoising, based on a self-supervised learning training strategy, specifically tailored to address the noise issues characterizing two-photon microscopy. To grapple with the dual challenges posed by the high noise levels and low signal strength that are hallmarks of such images, we take cues

from the cutting-edge UNet++architecture [57], integrating a multi-scale network design to adeptly handle intricate textures and details essential for high-quality biological image analysis. This enables us to extract features at different scales while preserving image structure and effectively removing noise. Given the stronger occurrence of shot noise in regions with intensified signals, it is imperative to circumvent the pitfall of over-denoising, which can obliterate critical image details. This balance is especially crucial for preserving the integrity of subtle yet significant structures, such as neuronal synapses, which are vital for accurate biological interpretation. To effectively safeguard these essential elements, we have augmented our network with an attention mechanism. This attention module is adept at differentially weighting the salient structural features within the image, bolstering their retention during the denoising process. Figure 1 provides an overview of the image denoising method of UNet-Att. Image data preprocessing was firstly performed to divide the data into input and target images for training the UNet-Att network. The trained model was utilized for the ultimate image denoising tasks.

Network architecture

In order to learn the structural details from the raw data without clean counterpart images, image preprocessing proceeds before the model training. In the image preprocessing, our methodology involves the dissection of a temporal sequence of 3D image stacks into two distinct sets, each comprising frames in close temporal succession. As depicted in Fig. 2a and b, the 3D stack consisting of odd frames is used as the input data of the network, and the 3D stack consisting of even frames is used as the learning target data of the network. Two-photon microscopic images are typically affected by two predominant and distinct types of noise: Gaussian noise and scattering noise. By leveraging this inherent noise independence, our proposed self-supervised approach is adept at training models to denoise images without the need for pristine ground truth data. This strategy holds promise for enhancing the clarity and usability of two-photon microscopy images, which is a significant step forward in biomedical imaging analysis.

The overall architecture of UNet-Att is depicted in Fig. 2c and adopts a U-shaped structure with multiple dimensions. It comprises three key components: the downsampling module, the attention module, and the multidimensional upsampling module. The downsampling module is responsible for feature extraction at various scales and reducing the image resolution. The attention module scores the information of the shallow-layer pixels based on the deep-layer network's feature information, enabling the shallow-layer network to better represent image structure features through the skip connections. Lastly, the multidimensional upsampling module fuses the image information from both the deep and

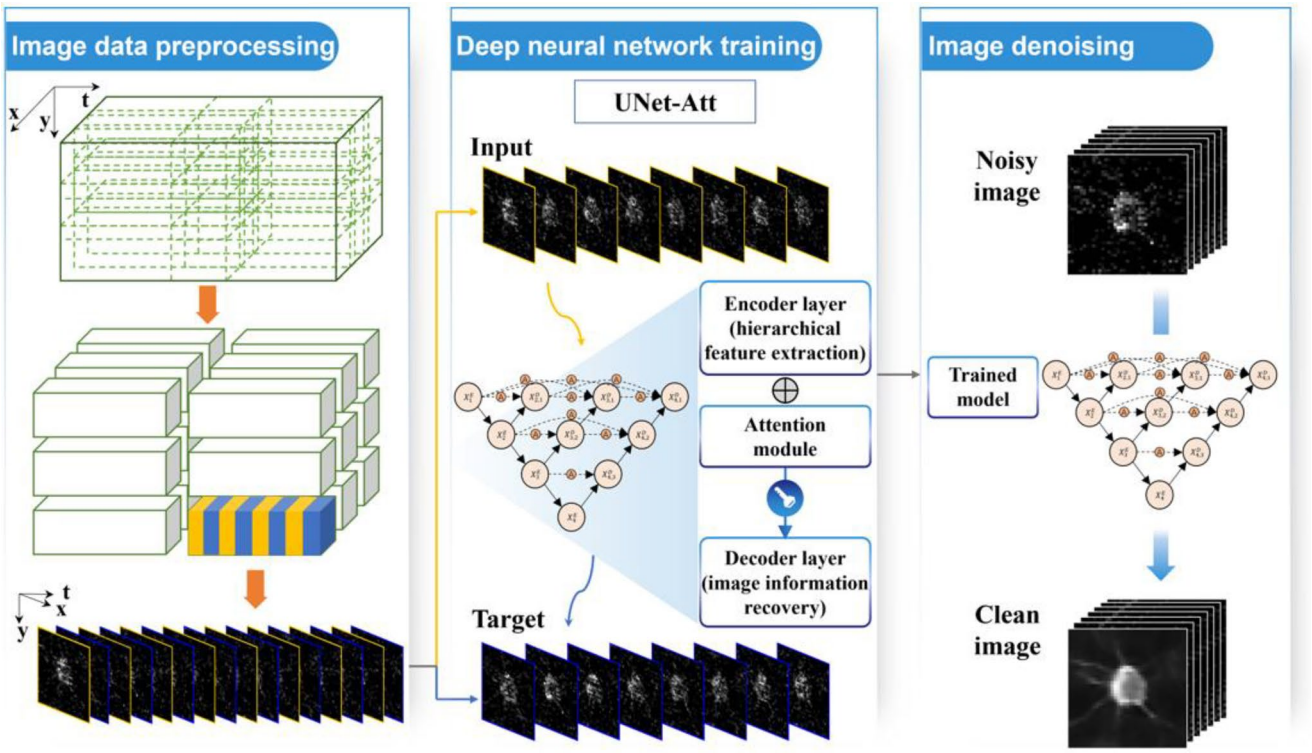

Fig. 1 Overview of the image denoising method of UNet-Att

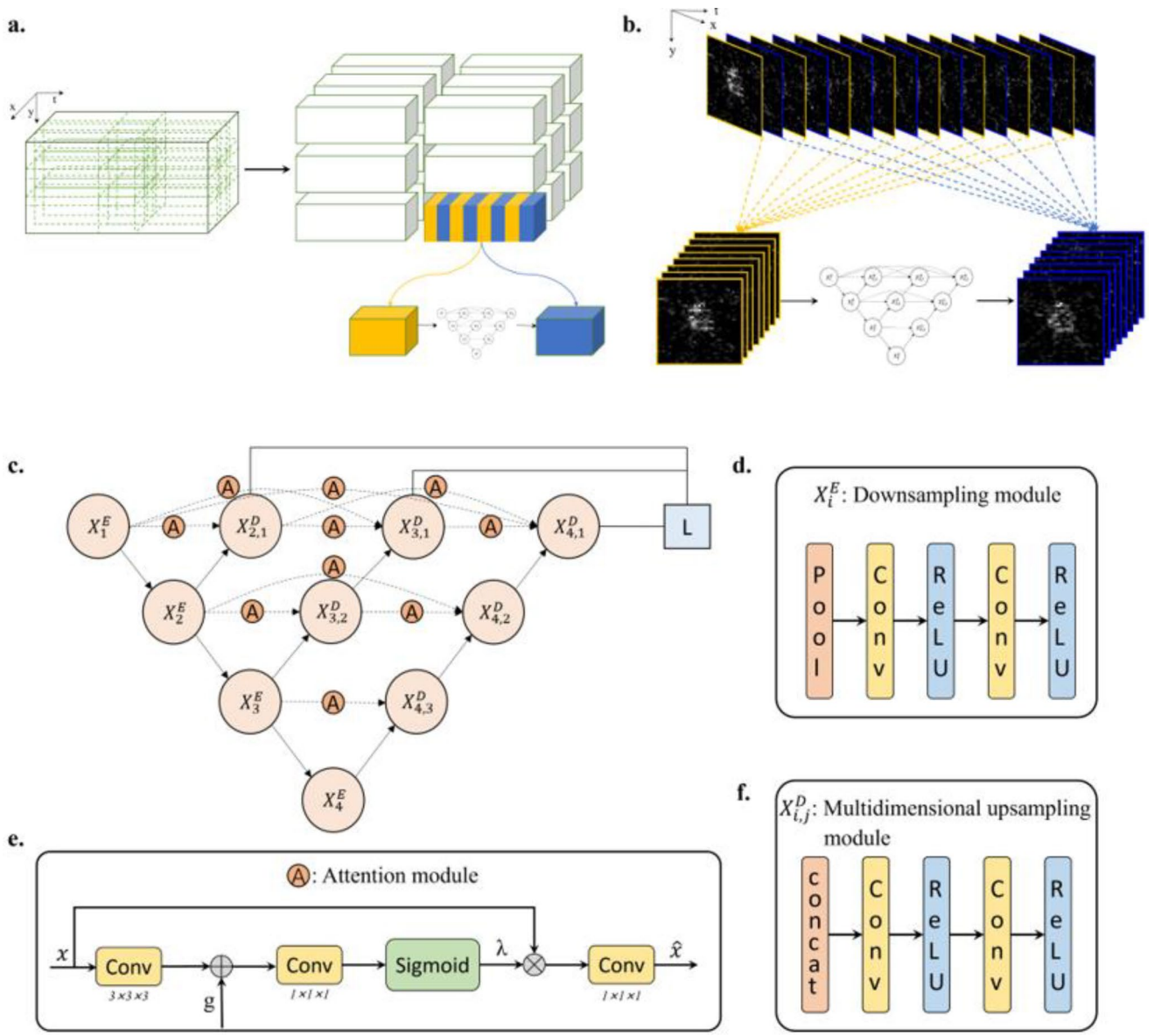

Fig. 2 The network architecture of UNet-Att and the structure of each module. **a**. Data preprocessing. The 3D data of size (6000,512,512) is sliced into multiple 3D data of size (512,64,64) and divides the network input data and the network learning target data. **b.** Delineation of network input data and target data. The 3D stack consisting of odd frames is used as the input data of the network, and the 3D stack consisting of even frames is used as the learning target data of the network. **c**. The overall network architecture of UNet-Att. x_i^E is the downsampling module, $x_{i,j}^D$ is the multi-dimensional upsampling module, A is the attention module, and L is the loss function. (**d**) Downsampling module, where Pool is the pooling layer, Conv is the convolutional layer, and ReLU is the activation layer. (**e**) Attention module, where x is the input to the network, g is the reference base of the network, Conv is the convolutional layer, Sigmoid is the normalization function, and $\widehat{x}$ is the network output. (**f**) Multidimensional upsampling module, where concat is the channel connection layer, Conv is the convolutional layer, and ReLU is the activation layer

shallow networks, as well as the output image information from the low-dimensional network, thereby restoring the image resolution.

The network structure of the downsampling module is illustrated in Fig. 2d, which comprises a pooling layer, two convolutional layers, and an activation layer. The convolutional layers are responsible for extracting image features, while the pooling layer downsamples the image data, enabling the model to focus on image information at different scales.

The input data consists of 3D data over time, represented by a matrix of size [t, H, W] (where t represents the number of time frames, H signifies the height of each image frame, and W denotes the width of each image frame). After applying the pooling layer, the image data is transformed into a matrix of size [t/2, H/2, W/2]. We utilize maximum pooling

to emphasize the contours of the image information. In the first convolutional layer, feature extraction is performed on the pooled image data, maintaining the same number of input and output channels. This design allows the network to learn features within the same channel dimension. In the second convolutional layer, the number of input and output channels differs, with the output channels being twice the number of input channels. This expansion in the feature learning dimension facilitates the network's ability to learn additional dimensions of image information. The activation layer plays a crucial role in applying a non-linear mapping to the output of the convolutional layer, enhancing the representational capability of the neural network. In this case, we employ the Rectified Linear Unit (ReLU) activation function, which accelerates network computation, mitigates the occurrence of overfitting, and alleviates the vanishing gradient problem.

The network structure of the attention module is depicted in Fig. 2e. The concept of attention mechanism is selectively focusing on specific data during the processing phase, rather than treating all data equally. This "focus" mechanism plays a crucial role in attention. In the context of image denoising, preserving the original information while reducing noise is of utmost importance. Hence, we introduce the idea of the attention mechanism in the denoising process of two-photon microscopy images.

To ensure the retention of overall structural information in the image after connecting the image information, we incorporate the attention module within the skip connection. Specifically, we assign the corresponding up-sampling module layer, denoted as x, and the output of the preceding layer of the neural network, denoted as g, to different convolutional layer operations. Here, x serves as the input to the attention module, while g acts as the reference for the attention module. By leveraging the overall structural information derived from the deeper network, the attention module applies varying degrees of attention to each pixel of x. The values are normalized using the Sigmoid function, with larger values of λ indicating a higher degree of attention assigned to a particular region. Then this attention coefficient is multiplied by x, resulting in the attention-processed image. Finally, the image information is integrated through a $1\times1\times1$ convolutional layer to obtain the attention module-processed image, $\widehat{x}$.

The fundamental network structure of the multidimensional upsampling module is illustrated in Fig. 2f. After the network extracts image features at different scales through downsampling, upsampling is required to restore the image resolution to its original form. Simply upsampling the output layer of the network can result in missing and distorted image information. To address this, we employ a skip connection, integrating the output of the corresponding encoding layer with the output of the previous network layer. In this process, we pass the output of the corresponding encoding layer through the attention module, which applies attention processing to the image information within the encoding layer. Subsequently, we perform information stitching with the output of the upper layer of the network, enabling the preservation of more image texture information. The skip connections within our multidimensional upsampling module not only encompass the connections between the encoding and decoding layers of the U-shaped network in the same dimension but also involve connections between U-shaped networks of different dimensions. When establishing a skip connection within a higher-dimensional U-shaped network, the corresponding decoding layer of the lower-dimensional U-shaped network is passed through the attention module, processed by attention, and then stitched with the corresponding layer of the respective dimensional U-shaped network, as depicted in Fig. 2. After performing the information stitching between the corresponding encoding layer of the same dimension, the decoding layer of the preceding layer, and all the corresponding decoding layers of lower dimensions, the information is fused through a series of convolutional layers to learn image features, reduce the number of channels to match the output channels of the feature map, and combine information from different channels. Additionally, another convolutional layer is applied with a consistent number of input and output channels to enhance the expressive capabilities of the network.

At the conclusion of the decoding layer within each dimension, the output of each dimensional network undergoes a $1\times1\times1$ convolutional layer. This step facilitates the integration of the results learned by the network across channels, generating an output of the same size as the network input. Finally, we average the outputs of the U-shaped networks across different dimensions to obtain the overall network output.

Loss function

The loss function we use is the arithmetic mean of the L_1 parametric loss and L_2 parametric loss with,

$$L_1 = \sum_{i=1}^{n} |y_i - f(x_i)| \tag{2}$$

$$L_2 = \sum_{i=1}^{n} (y_i - f(x_i))^2 \tag{3}$$

where y_i denotes the target pixel value and $f(x_i)$ denotes the estimated pixel value of the network. L_1 parametric tends to minimize the absolute difference between the pixel value of the target image and the pixel value of the estimated image, outliers do not cause particularly large losses,

fluctuations are small and stable, so it is more robust. L_2 parametric tends to minimize the squared error between the two, outliers have a large impact on the loss, so it makes the image smoother.

Experiments and results

To evaluate the performance of the proposed UNet-Att network, the architecture was trained using a sequence of two-photon micrographs depicting calcium activity in mouse brain cells. The dataset was divided into input and target images for network training. Using the trained model, denoising experiments were conducted on additional two-photon microscopy images with low signal-to-noise ratios (SNR) from the same category. The effectiveness of the denoising was assessed using evaluation metrics such as Peak Signal-to-Noise Ratio (PSNR), Structural Similarity Index Measure (SSIM), and Normalized Root Mean Square Error (NRMSE). A systematic comparison study was also conducted with other leading denoising methods to benchmark the performance of the UNet-Att network.

Datasets

In two-photon microscopy applications, it is common to observe live cells over extended periods, necessitating a low-illumination experimental environment and rapid imaging by the microscope. Consequently, challenging data is frequently produced, characterized by temporal stacks of images and high levels of noise. As a result, complex data is frequently produced, including temporal stacks of images and significant levels of noise.

For our research, we utilize a dataset consisting of two-photon micrographs from a three-dimensional data series captured over time. The current training set is derived from the publicly available dataset introduced by Li et al. [36]. Specifically, it consists of a sequence of two-photon micrographs depicting calcium activity in mouse brain cells. The observed field of view spans 550×575 μm, with imaging conducted at depths ranging from 80 μm to 210 μm, encompassing cell bodies and dendrites. The utilized dataset comprised a total of 7 three-dimensional images, each containing 6000 frames with a resolution of 512×512 pixels and an interval of 0.02 s per frame. More specifically, 85% of these images were designated as the training dataset, while the remaining 15% were allocated for evaluation purposes as the test dataset.

Data preprocessing

The high resolution of our dataset presents a challenge when training the neural network directly. Therefore, it is necessary to preprocess the image data by dividing it into smaller sub-stacks. This approach allows for more efficient feeding of the data into the neural network and facilitates iterative parameter updates. To enhance the transition information within the data, we incorporate overlapping regions in all three dimensions during the subdivision process, as depicted in Fig. 2a.

After obtaining the sub-stacks, we adopt a specific data selection strategy for network learning, as illustrated in Fig. 2b. In our approach, the odd frames of the sub-stack are chosen as the input data, while the even frames serve as the target data. This selection result is designed to leverage deep learning self-supervision for model training. Our self-supervised strategy operates on the principle that training with noisy target images yields noisy gradients. However, the average gradient across the entire training set approximates the true gradient, which is akin to the noise-free image. Since the time interval between consecutive images is only 0.02 s, we can consider them as two sets of data with distinct noise distributions captured simultaneously. As the noise in two-photon micrographs images is independently distributed around the actual signal, the average value of pixels over a certain time period tends to converge towards the true signal. Consequently, the average signal across the entire training set represents the real signal in its noise-free form, enabling the network to learn the characteristics of the true signal present in the training set.

Evaluation indicators

Noise reduction performance is commonly evaluated using two main categories of metrics: pixel-based statistics and structural feature similarity. In the pixel-based statistics category, we employed two widely used metrics are Mean Square Error (MSE) and Peak Signal-to-Noise Ratio (PSNR) [58]. On the other hand, the Structural Similarity (SSIM) [59] was utilized as the structural feature similarity category metrics.

Since MSE values tend to be large and not directly interpretable, they cannot be used as a direct quantifier of image quality. To address this, the root mean square error (RMSE) is obtained by taking the square root of MSE, and then normalized to obtain the normalized root mean square error (NRMSE). NRMSE provides a more directly interpretable measure of image quality. Smaller NRMSE values indicate higher image quality.

PSNR represents the ratio of the maximum power of the signal to the power of the noise in the signal. It improves the

sensitivity of the evaluation results by calculating the mean square error between the denoised and noiseless images and then converting it to the logarithmic domain. PSNR can be computed using Eq. (6), where I_{max} represents the maximum pixel value of the image.

$$PSNR = 10log_{10}\frac{I_{max}}{MSE} \tag{4}$$

PSNR is used to measure the difference between the ground truth image and the noise reduction image, allowing for evaluation of the performance of the noise reduction algorithm. A higher PSNR value indicates a better noise reduction effect.

SSIM represents the degree of structural similarity between two images, mainly focusing on the similarity of edges and textures in the images, which is more in line with human visual perception. SSIM calculates the statistical characteristics within the sliding window through sliding window operation, estimates the brightness of the image based on the mean of the image within the sliding window, and estimates the contrast and structural information similarity of the image based on the standard deviation and covariance of the image within the sliding window. Then move the window in pixels until the entire image is traversed, and calculate the global structural similarity. Its expression is given as:

$$SSIM = [l(x,y)]^{\alpha} \cdot [c(x,y)]^{\beta} \cdot [s(x,y)]^{\lambda} \tag{5}$$

where $l(x,y)$is the brightness contrast function, $c(x,y)$is the contrast function, $s(x,y)$ is the structural information contrast function, where α、β and λ indicates the importance of three modules, with values greater than 0. The expressions for the three functions are:

$$l(x,y) = \frac{2\mu_x\mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1} \tag{6}$$

$$c(x,y) = \frac{2\sigma_x\sigma_y + C_2}{\sigma_x^2 + \sigma_y^2 + C_2} \tag{7}$$

$$s(x,y) = \frac{2\sigma_{xy} + C_3}{\sigma_x\sigma_y + C_3} \tag{8}$$

In Eqs. (6), (7), and (8), μ_x and μ_y represents the mean of the denoised image and the denoised image, respectively, σ_x and σ_y represents the standard deviation of the denoised image and the noiseless image, respectively, σ_{xy} represents the covariance of two images. C_1、C_2 and C_3 is a very small constant to avoid having a denominator of 0. The SSIM value ranges from 0 to 1, where a higher value indicates a higher similarity between the denoised and noiseless images, less image distortion, better preservation of information in the denoised images, and superior performance of the model.

In addition to the aforementioned metrics, the richness of image information can be assessed by calculating the image entropy [60]. When calculating, it is usually necessary to perform grayscale normalization on the image to ensure that the probability of each pixel value is between 0 and 1. Image entropy is similar to information entropy and measures the complexity and randomness of information within an image. A higher image entropy signifies a greater amount of information in the image, indicating more complexity and higher image quality.

Experimental results and analysis

After the network model is trained, we apply it to other two-photon microscopy images with low signal-to-noise ratios from the same category as the training set. The estimation results are depicted in Fig. 3. Before denoising, two-photon microscopy images are heavily affected by noise interference, resulting in poor image readability. However, after being processed by our algorithm, the image noise is effectively removed, the image information is recovered, and image quality and readability improve significantly.

To validate the denoising efficacy of our proposed algorithm, UNet-Att, on real noise images with low signal-to-noise ratios in two-photon microscopy, we compare its performance with classical image denoising algorithms and mainstream fluorescence microscopy image denoising algorithms from recent years. We evaluate the effectiveness of UNet-Att based on three key metrics: PSNR, SSIM, and NRMSE. Additionally, we carefully examine the denoised images produced by each algorithm to assess the enhancement effect of UNet-Att on image denoising by observing intricate image details. Furthermore, we conduct several ablation experiments to verify the effectiveness of each module within UNet-Att for the task of denoising two-photon microscopic images.

Comparison experiments

In comparison, we selected three algorithms: DnCNN, Noise2noise, and DeepCAD. These are classical deep learning denoising algorithms that have demonstrated notable performance in the field of fluorescence microscopy image denoising in recent years. To ensure fairness, we trained all algorithms using the same dataset and training parameters. Subsequently, we applied each algorithm's model to estimate the denoising of 3D two-photon microscopy images with low signal-to-noise ratios. The denoising results are

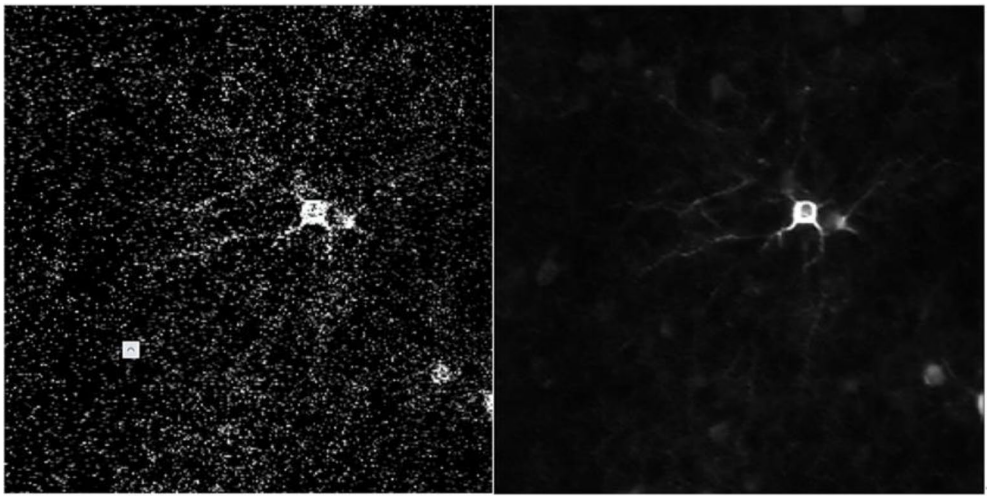

Fig. 3 Model denoising results. The left image is the original noisy image, and the right image is the image after denoising by the algorithm in this chapter

Table 2 Comparison of indicators of denoising results of different algorithms

Methods	PSNR/dB↑	SSIM↑	NRMSE↓
Raw	32.19	0.8553	0.5047
DnCNN [41]	37.52	0.9436	0.2737
Noise2Noise [38]	39.61	0.9591	0.2155
DeepCAD [36]	41.36	0.9679	0.1758
UNet-Att (ours)	**41.61**	**0.9684**	**0.1709**

presented in Table 2, which compares the metrics of the denoised images.

The results presented in Table 2 represent the mean values obtained from the algorithm models. Observing the PSNR metric, our algorithm produced the highest value, resulting in a 9.42dB increase in PSNR compared to the noise image. This demonstrates that our algorithm minimizes distortion and exhibits excellent noise removal capabilities. Similarly, our algorithm achieved the highest value in the SSIM metric, indicating a superior structural similarity to the noiseless image and preserving the image's texture and structure. The improvement in structural similarity amounts to 0.1131 when compared to the noisy image.

Furthermore, our algorithm attained the minimum value in the NRMSE metric, signifying the smallest pixel difference between our algorithm's output and the noiseless image, and reducing the degree of difference by 0.3338. Consequently, our algorithm outperforms mainstream algorithms for fluorescence microscopic image denoising, as demonstrated by the comparison between the denoised images and the noise-free images. Additionally, we computed the image entropy of the denoised images to assess the information richness achieved by different algorithms, as depicted in Fig. 4. It is evident that our algorithm effectively restores the image's information richness to an optimal level through the denoising process.

Furthermore, we conducted a visual comparison between the denoised images obtained by different algorithms and the noise-free images, which were obtained by averaging high signal-to-noise images. Figure 5 illustrates the results of this comparison. Prior to the denoising process, the two-photon microscopy images suffer from significant noise interference, obscuring crucial information. The presence of shot noise not only affects the background with a certain level of noise but also introduces higher noise levels to the calcium material within the cells of interest, resulting in poor image readability.

However, after applying the image denoising algorithms, notable differences can be observed. DnCNN exhibits weak denoising capabilities, as residual noise remains in the image. Conversely, the other three algorithms successfully remove a substantial portion of the image noise. Regarding the preservation of detailed information, such as the texture structure of the image, Noise2noise and DeepCAD exhibit inconspicuous and dark pixels in neuronal synapses, and encounter challenges like over-smoothing when handling slender neuronal synapses. In contrast, our algorithm, UNet-Att, effectively recovers the brightness information of neuronal synapses with higher clarity. When examining the processing of neuronal cells, Noise2noise yields structurally incomplete results, while DeepCAD suffers from insufficient contrast. In comparison, our algorithm produces results that closely resemble the real image. This further validates the effectiveness of the UNet-Att algorithm, as it not only removes noise effectively but also preserves essential image information, resulting in images that closely resemble the ground truth image.

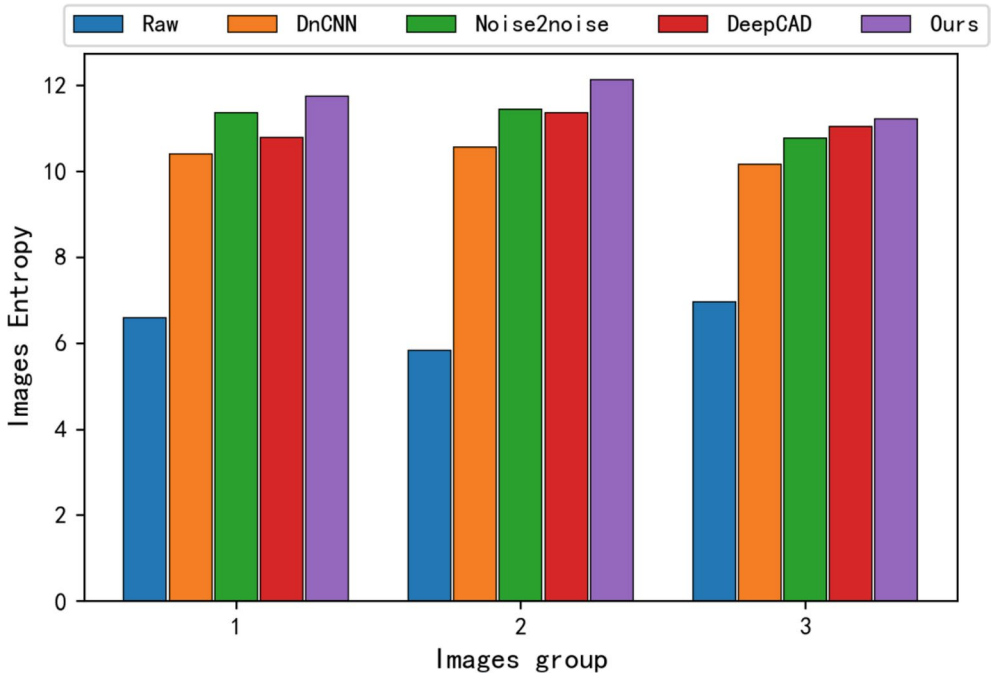

Fig. 4 Comparison of image entropy of denoising results of different algorithms. The vertical coordinate is the image entropy value of the denoised image and the horizontal coordinate is the different image groups

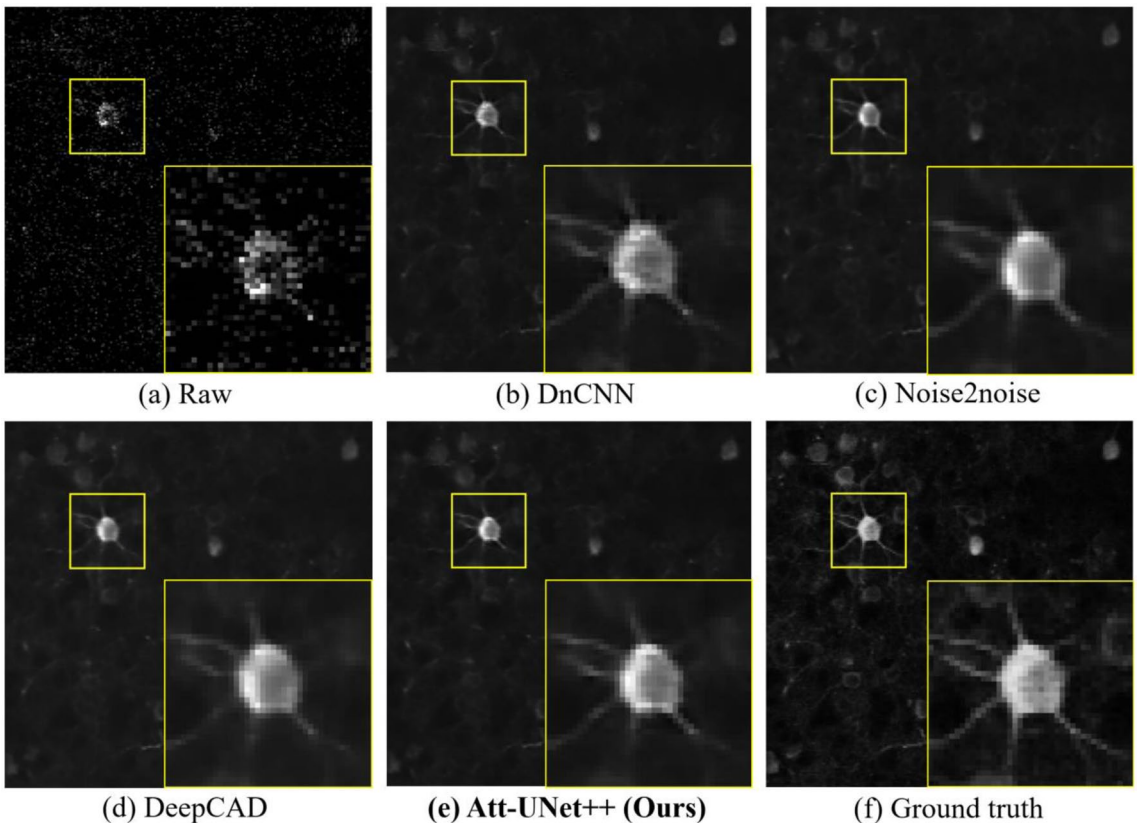

Fig. 5 Visualization comparison of this algorithm with other algorithms. **a** shows the original noisy image; **b** shows the denoised image of the comparison method DnCNN; **c** shows the denoised image of the comparison method Noise2noise; **d** shows the denoised image of the comparison method DeepCAD; **e** shows the denoised image of the algorithm in this chapter; **f** shows the corresponding noise-free image

Ablation experiments

To investigate the role of each module in the two-photon microscopic image denoising task, ablation experiments were conducted on the algorithm proposed in this paper. Ablation experiments involve removing or modifying specific components of the system to assess their impact on the overall performance. In this paper, the ablation experiments focused on three variations of the network structure:

1) The attention module was removed from the network architecture, resulting in a UNet++type network. This network configuration solely employed the downsampling and multidimensional upsampling modules. The resulting network structure is depicted in Fig. 6a.
2) The multidimensional structure in the multidimensional upsampling module was removed, resulting in a network architecture biased towards a unidimensional U-shaped network. This variant is represented Attention-Unet, and the network structure is shown in Fig. 6b.
3) A partial arrangement of the attention module was reduced by removing the attention module from the network layer of the multidimensional upsampling module. This variant, referred to as UNet-Att without DA. The network structure is depicted in Fig. 6c.

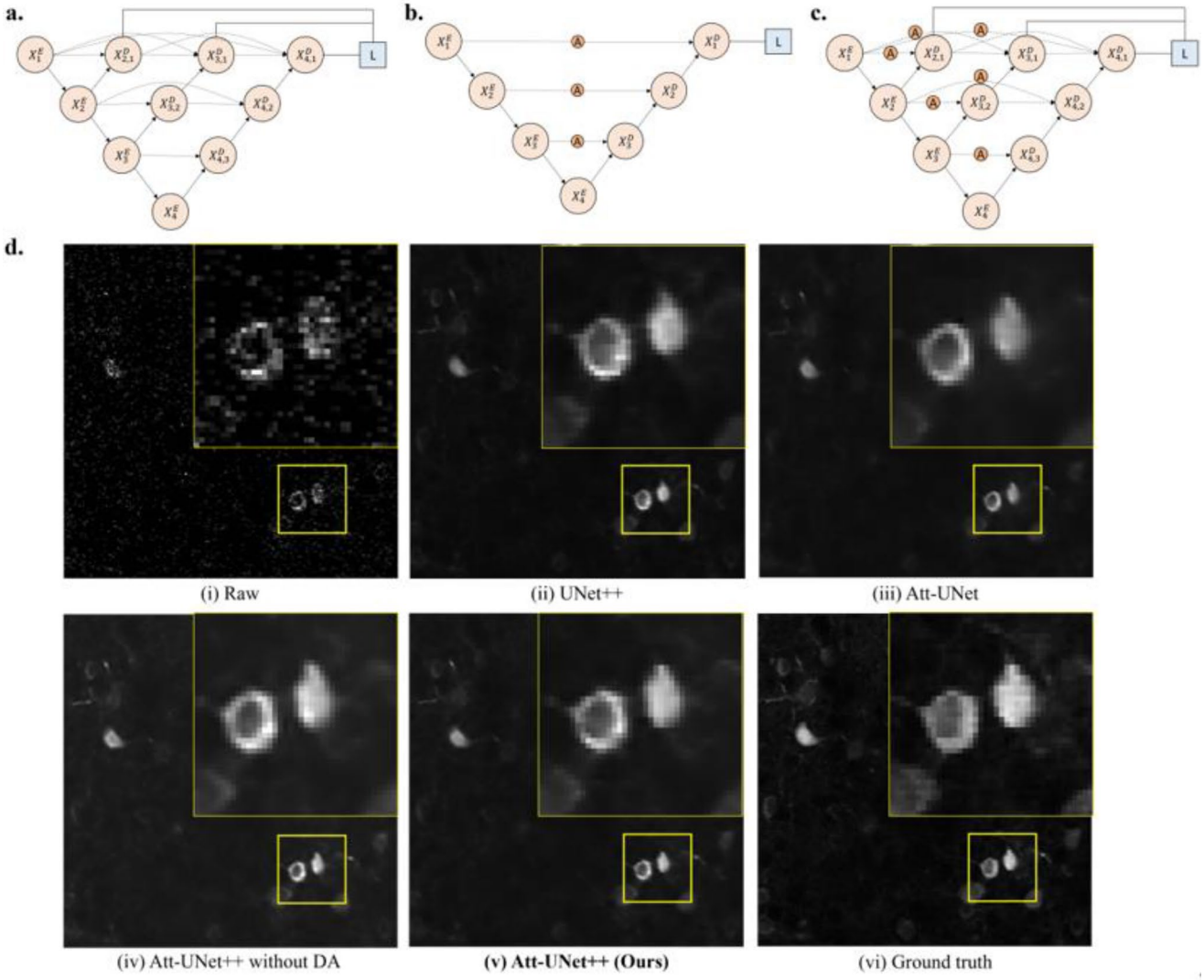

Fig. 6 Structure of each network of the ablation experiment and visual comparison of denoising results of ablation experiments. **a** shows the structure of the ablation network after removing the attention module; **b** shows the structure of the ablation network after removing the multidimensional structure in the multidimensional upsampling; **c** shows the structure of the ablation network after removing the attention module between the upsampling layers in different dimensions; **d** shows visual comparison of denoising results of ablation experiments: **(i)** is the original noisy image; **(ii)** is the denoised image of the ablation network UNet++; **(iii)** is the denoised image of the ablation network Att-UNet; **(iv)** is the denoised image of the ablation network UNet-Att without DA; **(v)** is the denoised image of the algorithm in this chapter image; **(vi)** is the corresponding noise-free image

Table 3 Comparison of indicators of denoising results of ablation experiments

Methods	PSNR/dB↑	SSIM↑	NRMSE↓
Raw	32.19	0.8553	0.5047
UNet++	40.73	0.9643	0.1910
Attention-Unet	38.37	0.9484	0.2503
UNet-Att without DA	38.63	0.9553	0.2415
UNet-Att	**41.61**	**0.9684**	**0.1709**

The results of the ablation experiments, presented as mean metrics across algorithmic iterations, are summarized in Table 3. The comparative data unambiguously demonstrate that the denoising capabilities of UNet-Att surpass its counterparts across all three evaluative criteria: Peak Signal-to-Noise Ratio (PSNR), Structural Similarity Index Measure (SSIM), and Normalized Root Mean Square Error (NRMSE). Consequently, employing the attention module and the multidimensional upsampling module both contribute to improving the efficacy of two-photon microscopy image denoising. Furthermore, the overall arrangement of the attention module proves great advantage in enhancing the denoising performance.

In addition, a visual comparison was conducted to assess the result of the ablation experiments, as illustrated in Fig. 6d. The results reveal significant insights into the effectiveness of each module. Notably, removing the attention module leads to a noticeable loss of information, indicating its crucial role in preserving image details. Similarly, the removal of the multidimensional upsampling module also results in substantial information loss, which is visually evident. Moreover, eliminating the attention module arrangement in the upsampling network layer reduces the denoising ability of the algorithm, allowing residual noise to persist in the image. Consequently, it can be concluded that the careful configuration and placement of each module in our UNet-Att algorithm are instrumental in achieving effective image denoising and preserving the essential information. These design choices significantly enhance the denoising performance of two-photon microscopic images.

Conclusion and discussion

The quantitative analysis of cellular and tissue imagery is of pivotal for elucidating their functional dynamics [61–63]. This paper presents a novel denoising algorithm, UNet-Att, specifically designed to address the denoising challenges encountered in two-photon microscopic images. By combining the concepts of the UNet++ network and attention mechanism, our proposed algorithm effectively suppresses noise while preserving the essential texture and structural information in fluorescent microscopic images.

UNet-Att harnesses the power of self-supervision and leverages an attention module to prioritize structural integrity, combined with multi-scale feature extraction capabilities inherent in the UNet++ framework. It employs a downsampling module to distill multi-scale image features, an attention module to accentuate structural information, and a multi-dimensional upsampling module to reconstruct image details with precision. We validate our algorithm on three-dimensional two-photon microscopy images characterized by a low signal-to-noise ratio, successfully eliminating image noise while maintaining the structural information.

Through exhaustive comparative evaluations, our UNet-Att algorithm demonstrates superior denoising efficacy relative to prevailing methods utilized in fluorescence microscopy image denoising. Ablation studies further validate the indispensability of each integrated module within our algorithm. While image analysis remains a cornerstone of our study, we acknowledge the parallel importance of theoretical modeling in understanding gene/protein signaling network dynamics. By a combination of our image analysis with network dynamics modeling, we aim to a more profound comprehension of regulatory mechanisms, potentially unveiling novel therapeutic targets for diseases intervention.

The proposed method UNet-Att gives priority to the denoising accuracy. The network architecture of the denoising algorithm in this study is more complicated and intricate, resulting in a trade-off for increased training time. However, after the training process, the denoising process didn't take more time compared with the mainstream methods. The ability to obtain clearer images with preserved details supports cutting-edge research in neuroscience, cell biology, and other biomedical fields. This can lead to new discoveries and advancements in understanding complex biological systems. The denoising efficiency would be further enhanced with the combination with diffusion method, transformer, and large language model (LLM) for future relevant exploration work. Recently the advanced diffractive neural network relying on stacked intelligent metasurfaces (sim) has been proposed for image reconstruction [64]. The sim-based diffractive neural network can perform edge detection and image enhancement directly in the optical domain, which could speed up preprocessing steps in computer vision pipelines. The integration of such a network holds the potential in future development of high-speed image denoising methods.

Acknowledgements The dataset for this work was derived from the publicly available data of Li et al. for which special thanks are given.

Funding This work is supported by the Ministry of Science and Technology of the People's Republic of China (STI2030-Major Proj-

ects2021ZD0201900), the National Natural Science Foundation of China under Grant 12090052, Foundation of Education Department of Liaoning Province (Grant No. LJKZ0280), the Fundamental Research Funds for the Central Universities (Grant No. 20720230017), and the Wenzhou Institute, University of Chinese Academy of Sciences' start-up fund (Grant No. WIUCASQD2023007).

Data availability The raw data was derived from the publicaly shared data available at https://github.com/cabooster/DeepCAD/tree/master/dataset. UNet-Att has been packaged as a python package. We shared all the code of UNet-Att in our GitHub repository https://github.com/ZjjDh/UNet-Att.

Declarations

Conflict of interest The authors declare that they have no conflicts of interest.

References

1. Denk W, Delaney KR, Gelperin A, Kleinfeld D, Strowbridge BW, Tank DW, Yuste R (1994) Anatomical and functional imaging of neurons usinG 2-photon laser-scanning microscopy. J Neurosci Methods 54(2):151–162
2. Helmchen F, Denk W (2005) Deep tissue two-photon microscopy. Nat Methods 2(12):932–940
3. Grienberger C, Konnerth A (2012) Imaging calcium in neurons. Neuron 73(5):862–885
4. Webb RH (1996) Confocal optical microscopy. Rep Prog Phys 59(3):427–471
5. Wang S, Lin B, Lin G, Lin R, Huang F, Liu W, Wang X, Liu X, Zhang Y, Wang F, Lin Y, Chen L, Chen J (2020) Automated label-free detection of injured neuron with deep learning by two-photon microscopy. J Biophotonics 13(1):1–13
6. Soltanian-Zadeh S, Sahingur K, Blau S, Gong Y, Farsiu S (2019) Fast and robust active neuron segmentation in two-photon calcium imaging using spatiotemporal deep learning. Proc Natl Acad Sci USA 116(17):8554–8563
7. Gu S, Zhang L, Zuo W, Feng X (2014) In *Weighted nuclear norm minimization with application to image denoising*, Proceedings of the IEEE conference on computer vision and pattern recognition, pp 2862–2869
8. Chen Y, Pock T (2016) Trainable nonlinear reaction diffusion: a flexible framework for fast and effective image restoration. 39(6):1256–1272
9. Zoran D, Weiss Y (2011) In *From learning models of natural image patches to whole image restoration*, international conference on computer vision, IEEE: 2011; pp 479–486
10. Buades A, Coll B, Morel J-M (2005) In *A non-local algorithm for image denoising*, IEEE computer society conference on computer vision and pattern recognition (CVPR'05), Ieee: 2005; pp 60–65
11. Dabov K, Foi A, Katkovnik V, Egiazarian K (2007) Image denoising by sparse 3-D transform-domain collaborative filtering. 16(8):2080–2095
12. Razmjooy N, Ramezani M, Ghadimi N (2017) Imperialist competitive algorithm-based optimization of neuro-fuzzy system parameters for automatic red-eye removal. 19:1144–1156
13. Elad M, Aharon M (2006) Image denoising via sparse and redundant representations over learned dictionaries. IEEE Trans Image Process 15(12):3736–3745
14. Meiniel W, Olivo-Marin J-C, Angelini E (2018) Denoising of microscopy images: a review of the state-of-the-art, and a new sparsity-based method. 27(8):3842–3856
15. Zhang L, Zhang J, Gao W, Bai F, Li N, Ghadimi N (2024) A deep learning outline aimed at prompt skin cancer detection utilizing gated recurrent unit networks and improved orca predation algorithm. 90:105858
16. Liu H, Ghadimi N (2024) Hybrid convolutional neural network and flexible dwarf mongoose optimization algorithm for strong kidney stone diagnosis. 91:106024
17. Cai X, Li X, Razmjooy N, Ghadimi N (2021) Breast cancer diagnosis by convolutional neural network and advanced thermal exchange optimization algorithm. Computational and Mathematical Methods in Medicine. 2021
18. Xu Z, Sheykhahmad FR, Ghadimi N, Razmjooy N (2020) Computer-aided diagnosis of skin cancer based on soft computing techniques. 15(1):860–871
19. Razmjooy N, Sheykhahmad FR, Ghadimi N (2018) A hybrid neural network–world cup optimization algorithm for melanoma detection. 13(1):9–16
20. He Q, Zhong C-Q, Li X, Guo H, Li Y, Gao M, Yu R, Liu X, Zhang F, Guo D, Ye F, Guo T, Shuai J, Han J (2023) Dear-DIAXMBD: deep autoencoder enables deconvolution of Data-Independent Acquisition Proteomics. Res (Washington D C) 6:1–14
21. Qian X, Qiu Y, He Q, Lu Y, Lin H, Xu F, Zhu F, Liu Z, Li X, Cao Y, Shuai J (2021) A review of methods for Sleep Arousal Detection using Polysomnographic signals. Brain Sci 11(10):1–27
22. Hu H, Feng Z, Lin H, Zhao J, Zhang Y, Xu F, Chen L, Chen F, Ma Y, Su J, Zhao Q, Shuai J (2023) Modeling and analyzing single-cell multimodal data with deep parametric inference. Brief Bioinform 24(1):1–13
23. Zhao J, Sun J, Shuai SC, Zhao Q, Shuai J (2022) Predicting potential interactions between lncRNAs and proteins via combined graph auto-encoder methods. Brief Bioinform 24(1):1–9
24. Wang W, Zhang L, Sun J, Zhao Q, Shuai J (2022) Predicting the potential human lncRNA-miRNA interactions based on graph convolution network with conditional random field. Brief Bioinform 23(6):1–9
25. LeCun Y, Bengio Y, Hinton G (2015) Deep learning. Nature 521(7553):436–444
26. Chiang Y-W, Sullivan B (1989) In *Multi-frame image restoration using a neural network*, Proceedings of the 32nd Midwest Symposium on Circuits and Systems, IEEE: pp 744–747
27. Jain V, Seung S (2008) Natural image denoising with convolutional networks. *21*
28. Xie J, Xu L, Chen E (2012) Image denoising and inpainting with deep neural networks. *25*
29. Burger HC, Schuler CJ, Harmeling S (2012) In *Image denoising: Can plain neural networks compete with BM3D?* IEEE conference on computer vision and pattern recognition, IEEE: 2012; pp 2392–2399

30. Goodfellow I, Bengio Y, Courville A (2016) *Deep learning*. MIT press
31. Mao X, Shen C, Yang Y-B (2016) Image restoration using very deep convolutional encoder-decoder networks with symmetric skip connections. *Advances in neural information processing systems 29*
32. Ronneberger O, Fischer P, Brox T (2015) *U-net: Convolutional networks for biomedical image segmentation*, Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5–9, 2015, proceedings, part III 18, Springer: pp 234–241
33. Siddique N, Paheding S, Elkin CP, Devabhaktuni V (2021) U-net and its variants for medical image segmentation: a review of theory and applications. IEEE Access 9:82031–82057
34. Zhang Y, Yu H (2018) Convolutional neural network based metal artifact reduction in X-ray computed tomography. IEEE Trans Med Imaging 37(6):1370–1381
35. Xiao T, Xu Y, Yang K, Zhang J, Peng Y, Zhang Z (2015) In *The application of two-level attention models in deep convolutional neural network for fine-grained image classification*, Proceedings of the IEEE conference on computer vision and pattern recognition, pp 842–850
36. Li X, Zhang G, Wu J, Zhang Y, Zhao Z, Lin X, Qiao H, Xie H, Wang H, Fang L (2021) Reinforcing neuron extraction and spike inference in calcium imaging using deep self-supervised denoising. Nat Methods 18(11):1395–1400
37. Ma C, Tan W, He R, Yan B (2024) Pretraining a foundation model for generalizable fluorescence microscopy-based image restoration. Nat Methods 1–10
38. Lehtinen J, Munkberg J, Hasselgren J, Laine S, Karras T, Aittala M, Aila T (2018) Noise2Noise: Learning image restoration without clean data. *arXiv preprint arXiv:1803.04189*
39. Ioffe S, Szegedy CI (2015) *Batch normalization: Accelerating deep network training by reducing internal covariate shift*, International conference on machine learning, pmlr: pp 448–456
40. He K, Zhang X, Ren S, Sun J (2016) In *Deep residual learning for image recognition*, Proceedings of the IEEE conference on computer vision and pattern recognition, pp 770–778
41. 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* (7), 3142–3155
42. Weigert M, Schmidt U, Boothe T, Müller A, Dibrov A, Jain A, Wilhelm B, Schmidt D, Broaddus C, Culley S (2018) Content-aware image restoration: pushing the limits of fluorescence microscopy. Nat Methods 15(12):1090–1097
43. Lee S, Negishi M, Urakubo H, Kasai H, Ishii S (2020) Mu-net: Multi-scale U-net for two-photon microscopy image denoising and restoration. Neural Netw 125:92–103
44. Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, Courville A, Bengio Y (2020) Generative adversarial networks. Commun ACM 63(11):139–144
45. Gidaris S, Singh P, Komodakis N (2018) Unsupervised representation learning by predicting image rotations
46. Zhai X, Oliver A, Kolesnikov A, Beyer L (2019) In *S4l: Self-supervised semi-supervised learning*, Proceedings of the IEEE/CVF international conference on computer vision, pp 1476–1485
47. Jing L, Tian Y (2020) Self-supervised visual feature learning with deep neural networks: a survey. 43(11):4037–4058
48. Xu K, Ba J, Kiros R, Cho K, Courville A, Salakhudinov R, Zemel R, Bengio Y (2015) In *Show, attend and tell: Neural image caption generation with visual attention*, International conference on machine learning, PMLR: pp 2048–2057
49. Hu J, Shen L, Sun GI (2018) *Squeeze-and-excitation networks*, Proceedings of the IEEE conference on computer vision and pattern recognition, pp 7132–7141
50. Oktay O, Schlemper J, Folgoc LL, Lee M, Heinrich M, Misawa K, Mori K, McDonagh S, Hammerla NY, Kainz B (2018) Attention u-net: Learning where to look for the pancreas. *arXiv preprint arXiv:1804.03999*
51. Lequyer J, Philip R, Sharma A, Hsu W-H, Pelletier L (2022) A fast blind zero-shot denoiser. Nat Mach Intell 4(11):953–963
52. Li X, Li Y, Zhou Y, Wu J, Zhao Z, Fan J, Deng F, Wu Z, Xiao G, He J (2023) Real-time denoising enables high-sensitivity fluorescence time-lapse imaging beyond the shot-noise limit. Nat Biotechnol 41(2):282–292
53. Li X, Hu X, Chen X, Fan J, Zhao Z, Wu J, Wang H, Dai Q (2023) Spatial redundancy transformer for self-supervised fluorescence image denoising. Nat Comput Sci 3(12):1067–1080
54. Bondani M, Allevi A, Zambra G, Paris MGA, Andreoni A (2007) Sub-shot-noise photon-number correlation in a mesoscopic twin beam of light. Phys Rev A 76(1):1–5
55. Boyat A, Joshi BK, Ieee In *Image Denoising using Wavelet Transform and Median Filtering*, 4th Nirma University International Conference on Engineering (NUiCONE), Ahmedabad, INDIA, 2013 Nov 28–30; Ahmedabad, INDIA, 2013
56. Laine RF, Jacquemet G, Krull A (2021) Imaging in focus: an introduction to denoising bioimages in the era of deep learning. Int J Biochem Cell Biol 140(106077):1–9
57. Zhou Z, Siddiquee MMR, Tajbakhsh N, Liang J (2020) UNet plus plus: redesigning skip connections to exploit Multiscale features in image segmentation. IEEE Trans Med Imaging 39(6):1856–1867
58. Avcibas I, Sankur B, Sayood K (2002) Statistical evaluation of image quality measures. J Electron Imaging 11(2):206–223
59. Wang Z, Bovik AC, Sheikh HR, Simoncelli EP (2004) Image quality assessment: from error visibility to structural similarity. IEEE Trans Image Process 13(4):600–612
60. Shannon CE (1997) The mathematical theory of communication (reprinted). M D Comput 14(4):306–317
61. Zhong J, Song Z, Zhang L, Li X, He Q, Lu Y, Kariko S, Shaw P, Liu L, Ye F, Li L, Shuai J (2022) Assembly of Guanine Crystals as a low-polarizing Broadband Multilayer Reflector in a spider, Phoroncidia rubroargentea. ACS Appl Mater Interfaces 14(29):32982–32993
62. Li X, Zhang P, Yin Z, Xu F, Yang Z-H, Jin J, Qu J, Liu Z, Qi H, Yao C (2022) Caspase-1 and gasdermin D afford the optimal targets with distinct switching strategies in NLRP1b inflammasome-induced cell death. Research
63. Wang J, Chen F, Ma Y, Wang L, Fei Z, Shuai J, Tang X, Zhou Q, Qin J (2023) XBound-Former: toward Cross-scale Boundary modeling in transformers. IEEE Trans Med Imaging 42(6):1735–1745
64. Liu H, An J, Jia X, Lin S, Yao X, Gan L, Clerckx B, Yuen C, Bennis M, Debbah M (2024) Stacked Intelligent Metasurfaces for Wireless Sensing and Communication: Applications and Challenges