IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received December 24, 2020, accepted January 21, 2021, date of publication January 26, 2021, date of current version February 10, 2021.

*Digital Object Identifier 10.1109/ACCESS.2021.3054728*

# Denoising Across Data Acquisition Modalities for Mesoscopic Scale Optical Neuroimaging

**TIANFANG ZHU, YUTONG HAN, ANAN LI, JING YUAN, GONG HUI, AND YUE GUAN, (Member, IEEE)**
Huazhong University of Science and Technology, Wuhan, Hubei 430074, China

Corresponding author: Yue Guan (yguan@hust.edu.cn)

This work was supported by the National Natural Science Foundation of China under Grant 81827901, Grant 61890954, and Grant 61890950.

**ABSTRACT** The ever-evolving mesoscopic scale optical imaging systems facilitate the new discoveries in the neuroscience research. They excel at providing a complete view of the long projections of a single neuron across the whole brain. Although the preparation protocols and the optical imaging systems are rapidly evolving, there are still some noise and artifacts interfering downstream data processing and analysis due to the defects in the imaging system or during the sample preparation. A specific denoising procedure is usually developed for one optical imaging system as a post-processing measure. However, the development of the optical imaging systems usually follows in an incremental manner. It would be better to adapt the denoising model to the new optical imaging system than training a denoising model from scratch. In this paper, the proposed learning schema and practice learn a new denoising model for a new optical imaging system based on the existing denoise models and bootstrap a denoising model without the ground-truth denoising labels. We achieve this through a CycleGAN based model and the fact that the optical imaging systems usually produce images both from the signal area and the fixture area. The experiments show that our proposed procedure can provide a comparable denoising performance against other state-of-the-art denoising methods for several optical neuroimaging datasets.

**INDEX TERMS** CycleGAN, denoising, mesoscopic optical imaging.

## I. INTRODUCTION

The mesoscopic scale optical imaging systems can capture images of the neuronal fibers, which may expand across a large part of the brain [1]. In this paper, we focus on the image denoising for the micro optical sectioning tomography (MOST) and its followup works [2], [3]. The imaging process and the result is illustrated in Fig. 1. The imaging system has several options to use different optical imaging devices to take the image after a slice is sectioned away on the top of the sample, including the structured illumination (SI) imaging system [4], and the time delay integration (TDI) imaging sytem [5]. The brain is usually embedded in the fixture compound to maintain its shape and the image of the slice will contain both the fixture area and the signal area, which is the brain slice region. The MOST imaging result is a series of 2D slices that display the optical acquired images from a model animal brain and a 3D reconstructed dataset describes the neurons morphologically.

It can provide concrete morphological evidence about the spatial distribution of a specific neuron type across the whole brain and the projection pattern between brain regions and nuclei [6], [7]. Due to the sample preparation process and the optical imaging system defects, it is necessary to enhance the results from the imaging systems through a denoising procedure. The denoised images will benefit the downstream data processing and analysis for the neuroscience research, such as the soma detection, the neuron tracing, and the projection pattern recognition [8]. However, the denoised neuroimaging datasets are generally not available for the supervised learning based denoising models because of the intimidating cost of the human labeling. The fixed feature-based denoising methods may help here when the cleaned images are not available [9], [10], but the tuning of the model for a slightly modified system are prone to be difficult and the result may diverge from what the neuroscientists expect [11]. Moreover, the optical imaging system is usually under a fast

The associate editor coordinating the review of this manuscript and approving it for publication was Yu Zhang.

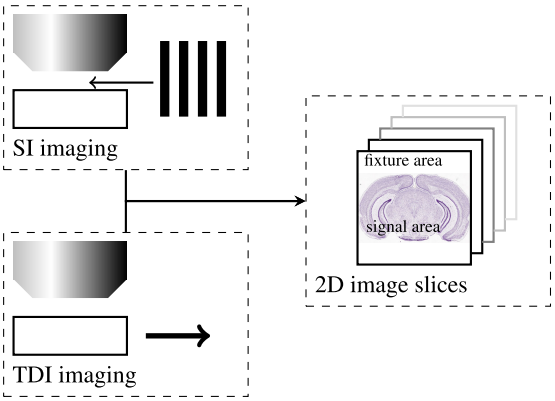



**FIGURE 1.** The SI and TDI optical imaging system used in MOST.

continuous development. This requires the denoising procedure to adapt to a new system without the manually labeled data if possible in order to provide a normalized dataset for the downstream analysis and to verify the new system's capacity. This is necessary even when the size of the available dataset is limited [12]–[14]. Moreover, the imaging result is also highly dependent on the labeling protocol used for the sample preparation. The staining methods are often related to the background noise level, the labeled neuron structure, and the brain vessel structure. The denoising procedures supporting different staining methods require an approach to transfer the already trained denoising model to a new model for the other staining method.

The image denoising has been studies for decades from different prespectives and is a key part of the MOST image processing pipeline. Classic methods start with applying varieties of image filters to the images to recover the pixels from their neighbours [15]. Following the image filters approach, more structured priors are constructed for the image denoising, either in the spatial space or the transform domain. These include nonlinear total variation based methods [16], collaborative filtering in transform-domain [17], and sparse reconstruction based methods [18]. As the booming deep learning research expands its territory in compute vision and natural language processing, denoising using deep neural network is intensively studied as well. One discussion bridge the classic method and the deep network based approach is shown in [9]. Another work uses networks with an extensive depth to denoise the image [19], [20]. The noise models beyond the Gaussian noise are also tested against the deep network model [21]. Dearling with more structured noise presented in the images is studied as well [22], [23]. In general, the supervised approach inevitably requires the available clean image set. On the other hand, the unsupervised deep denoising models usually exploit the relation within the noisy images itself and try to recover the underlying clean image. While denoising with deep learning mainly focuses on using the neural network to approximate the true clean images from the noisy ones and directly applying it to the MOST data may ignore some useful properties of the imaging process.

Although removing noise in the image seems to be a reduction process, image denoising is also an image reconstruction process where the noisy input image is the low quality input and the denoised one is the high quality recovered output. This is related to the generative model, such as the generative adversarial network (GAN) [24] and the conditional GAN [25]. In the GAN model, a generative model generates an instance in the target space, and a discriminative model tests the distance between the generated instances and the real samples [26]. There are works applying the GAN model for the medical imaging analysis [27], [28], and the image translation even without paired images using CycleGAN is studied as well [29]. Using the CycleGAN model is attractive in our application where the neuroscientists use multiple imaging system to observe the same brain tissue and they wish to convert the result images from one system to another to either verify the validation of the imaging system or use the downstream processing pipeline choice at their disposal. The CycleGAN model can convert images between different domains and fulfill these requirements from neuroscientists.

In this paper, we introduce an image denoising approach built for the MOST system relying on the optical imaging result and learn the models for the SI and TDI imaging systems using the CycleGAN approach. The contribution of the proposed work is in two folds, it includes:

1) The proposed model makes use of the signal area and fixture area, where both are usually captured simultaneously in one imaging process. It demonstrates the possibility to train a denoising model without in-domain high-quality images but using noise only images.
2) The experiments also shows it is possible to use a CycleGAN based denoising model for another optical imaging system based on an existing model using only the new noisy images.

The paper is organized as the following. We discuss the possibility of using fixture area to bootstrap a denoising model in section II-A, then lay out a CycleGAN based model to build a new denoising model using an existing one in section II-B. In the experiment section, we first test the denoising backbone against synthetic noisy images in section III-C. Then the bootstrap process with images from fixture area is performed and verified in section III-D. Finally, a CycleGAN based model is built for a new optical auqusition system in section III-E. The discussion about this paper and possible extension is in section IV.

## II. PROPOSED METHOD

Our proposed denoising solution consists of two parts. We first bootstrap a general denoising model based on the set of some clean natural images and a noisy version synthesized by combining the clean natural images and the images from the fixture area. The backbone of the denoising model is a modified DnUNet model [30]. Once a denoising model trained for one image acquisition modality is available, another

denoising model for the other optical system can be obtained through a CycleGAN learning strategy.

Sec. II-A introduces the bootstrap procedure for the denoising model training without the in-domain labels and Sec. II-B presents the training of a denosing model for another imaging modality by a CycleGAN approach.

### A. BOOTSTRAP THE DENOISING MODEL WITHOUT THE GROUND TRUTH

One difficulty in the optical neuroimaging denoising is that the high-quality ground truth images are not always available for training the denoising model. Meanwhile, the MOST system can capture the images of a brain tissue slice with both the fixture area and the signal area, as shown in Fig. 1. The part of the image outside the brain boundary and the areas without the brain structural signal inside the brain boundary will have the background additive noise only. Thus this set of images is used as the samples of the background noise. Here $\epsilon_i$, where $i = 1, \ldots, N_\epsilon$, is one of the cropped image from the area outside the brain boundary or without structural signal, $N_\epsilon$ is the number of noisy images available. Meanwhile we can crop $N_{im}$ images $I_{im,i}$, $i = 1, \ldots, N_{im}$ with the same size from the high quality clean image set. Then the noisy image $Y_i$ for the supervised denoising training can be an overlay of the noise image $\epsilon_i$ and $I_{im,i}$, as shown in Equ. (1)

$$Y_i = I_{im,i} + \alpha_b \epsilon_i, \tag{1}$$

where $\alpha_b$ is the noise strength coefficient. It may be necessary to truncate the $Y_i$ in the cases that the numerical summation overflows. The goal of this denoising model is to train the model $f_b(\cdot)$, such that $I_{im,i}$ is retrieved, i.e. $f_b(Y_i) \simeq I_{im,i}$. The clean image set in the experiment section for $I_{im,,i}$ is the natural scene images from [31].

The denoising backbone model is a modified UNet model [30] denoted as DnUNet. The model takes the noisy version $Y_i$ and try to recover the clean $I_{im,i}$. The structure of the model is illustrated in Fig. 2. Through the model, the input image undergoes a few max-poolings and upconvolutions to generate the denoised image though the multi-scale resolution steps. The depth of the first convolution layer is 64, it doubles after every max-pooling layer and reduces in half after every upconvolution layer. The batch normalization layer is inserted after the convolutional layer to provide a smoother training process. The activation function used in the network is leaky rectified linear unit (LReLU). The feature maps at the same resolution before the pooling layer directly concatenate with the feature maps immediate after the upconvolution layer. The original input is added to the network output to form the final result, this allows the network only learns to extract the noise from the input, rather than the content of the image. The mean-squared-error (MSE) is used as the cost function. This network structure with the multi-resolution and the skip-link allows a good feature representation while maintaining the original spatial resolution.

Although there will be discrepancy between the natural scene images and the optical neuron images, we try to improve the performance of $f_b(\cdot)$ based on DnUNet by tuning the model through a noise2noise process [32]. In this case, one signal image $I_{im,i}$ generates several noisy images $Y_{i,j}$ by adding to several pure noise image $\epsilon_j$ with the same weight $\alpha_b$. The DnUNet based model then takes a randomly chosen pair $(Y_{i,j}, Y_{i,j'})$ as the input and the target image. Due the fact that $E_j(Y_{i,j}) = I_{im,i}$ holds assuming $E_j\epsilon_j = 0$, the model is expected to output a sharper result than applying the DnUNet model directly over $(Y_i, I_{im,i})$.

### B. LEARN A DENOISING MODEL FOR OTHER IMAGING MODALITY

Although the denoising model obtained in Sec. II-A is often competent for the task on the imaging system it is trained for, it is also valuable if denoising model can adapt to a new imaging system assuming the underlying imaging target is the same. This is often the case in the optical neuroimaging systems, where the fluorescent signal is mainly determined by the target brain structure and the labeling strategy. While the difference in the optical imaging systems introduces a variety of noise pattern due to different imaging principals. The noise pattern may also be different because of an imaging system with a different configuration. All these varieties lead to the expectation of the denoising model can adapt to the new domain. While the transfer learning may provide a new denoising model for the new imaging system, neuroscientist may also want to see the target recognized by the two denoising system is the same and the conversion between the image output from these two imaging systems may provide some suggestion that the denoised images are indeed two aspects of the same brain tissue. Thus rather than obtaining the new denoising model by the transfer learning approach to fine tune the existing denoising model to adapt the new imaging system, using a CycleGAN structure allows us to have the ability to convert images from the two imaging systems to each other.

To adapt the denoising model to a new imaging system, we can use the CycleGAN based architecture, which transfers the style between the source domain and the target domain by forming a ring network with two symmetric GANs as shown in Fig. 3. While CycleGAN's greatest advantage is that it does not require paired data, it will only work well if the image content of the source domain is similar to that of the target domain. This is the case for the optical neuroimaging system in general, where the neuron structures are the image acqusition target and the imaging systems just capture its property from the different modalities. The diagram shown in Fig. 3 is an example of a quick migration of the denoising results from an SI system to a TDI system. Firstly, the DnUNet trained on the simulation noise dataset as described in section II-A is used to denoise the signal images of SI system, and a set of the denoised SI signal images is obtained. This process is denoted as the solid arrow line in Fig. 3. Then, the denoised SI signal images go through the DnUNet backbone to generate the noisy TDI style images; together with the original noisy TDI images, a discriminator is trained to form a GAN with

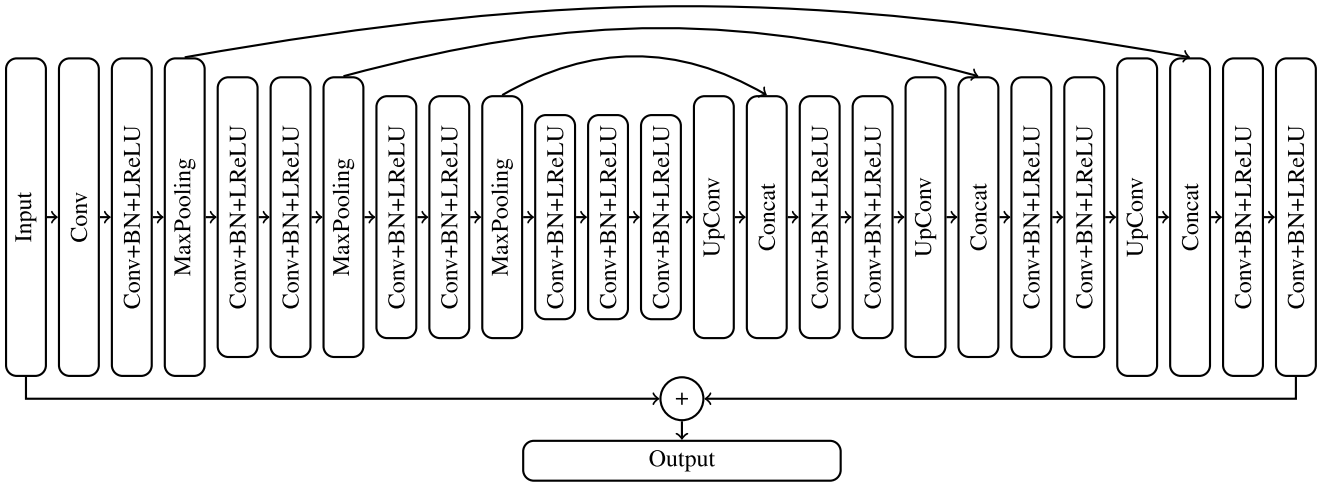



**FIGURE 2.** DnUNet denoising backbone model based on UNet.

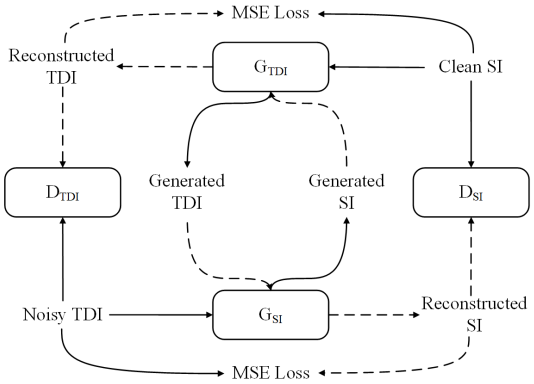



**FIGURE 3.** Learn a denoising model in TDI noise domain based on the denoising model for SI.

the DnUNet backbone. As now, both the generated noisy TDI images and its corresponding denoised images are available, thus a specific denoising model can be trained for the new TDI system. To close the CycleGAN loop, we also repeat the second step to build a GAN model to convert the denoised TDI images back to the noisy SI images. The object function to train the two GANs in one loop is shown in Equ. (2)

$$MSE(G_{SI}(G_{TDI}(SI_c)), SI_c) + MSE(G_{TDI}(G_{SI}(TDI_n)), TDI_n), \quad (2)$$

where $G_{SI}$ is the generative model to convert the noisy TDI images to the clean SI images, $G_{TDI}$ is the generative model to convert the clean SI images to the noisy TDI images. $SI_c$ is the clean SI image dataset and $TDI_n$ is the noisy TDI image dataset. $MSE(\cdot)$ is the mean-squared-error loss. This second round training is denoted as the dashed arrow line in Fig. 3.

The network used in building the GAN, denoted as UGAN is shown in Fig. 4. The generative network (G) for converting the clean image to the noisy image in the other modality is the same DnUNet used in previous section. The discriminator (D) part has a combination of the convolution layers and the last

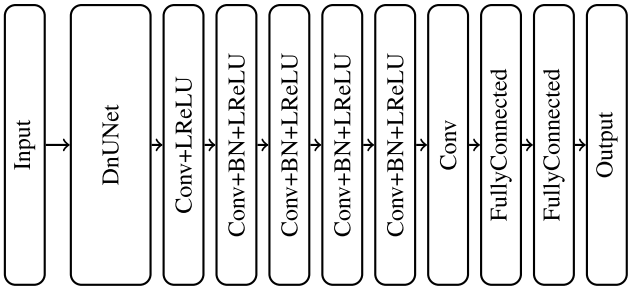



**FIGURE 4.** The GAN model to convert a clean image to another noisy image in the other modality.

two fully connected layers. The convolution layer uses both the batch normalization and the leaky ReLU activation unit. The first 3 convolution layers use 64 $3 \times 3$ filters, while the remaining 3 convolution layers use 128f $3 \times 3$ filters. The fully connected layers use the Tanh function as the activation function. The proposed GAN model is a mix-max game with the Wasserstein cost. The GAN model converting the denoised SI images to the noisy TDI images alike is shown in (3). The other GAN model is built alike in the denoised TDI images to the noisy SI images direction.

$$\min_G \max_D E_{TDI_n, G(SI_C)} W(D(G(SI_c)), TDI_n) \quad (3)$$

where $W(\cdot, \cdot)$ is the Wasserstein cost, $SI_c$ is the denoised images from SI imaging system, $TDI_n$ is the raw noisy images from the TDI imaging system. The cost is averaged across the training set.

## III. EXPERIMENT

In this experiment section, the experiment result for the model bootstrap step and the denoising GAN step of the proposed approach are presented and compare the denoising performance against other baseline methods. We first test the DnUNet model on a set of synthetic dataset and check the result visually to verify the capability of the model in

Sec. III-C. Then we present the noisy fixture area images from the SI imaging system, generate the image pairs for the DnUNet training and bootstrap the denoising model without the clean images in Sec. III-D. At last, the CycleGAN model is built to obtain a denoising model for the TDI system based on the denoising model for the SI system in Sec. III-E.

### A. DATA ACQUISITION

There are several images dataset used in the training and the prediction of the proposed denoising model. The image dataset for bootstrapping the denoising system is the natural scene dataset [31]. A total of 2688 images from the dataset is used in the training. The SI imaging system used in the MOST system serves as the baseline optical imaging system [33]. The data collected is from a PV-cre line mouse with an AAV GFP injection site in the M1 region. The imaging resolution is $0.32\mu m$ per pixel along both the x- and y-axis and $1\mu m$ per pixel along the z-axis. The imaging process takes around one week continuously. The whole SI image dataset contains around 5000 slices and the size of the raw data is around 10 terabytes. The target imaging system is the optical TDI system also used in the MOST system. The specification of the TDI system is similar to those found in the SI system. The imaging resolution of the TDI system is also $0.32\mu m$ per pixel along both the x- and y-axis and $1\mu m$ per pixel along the z-axis. $10\mu m$ slices are combined into one image by average and the combined images are used in the following experiment. As from the data source description above, the only high-quality images used in this work are the natural scene images.

### B. DATA PREPARATION

To capture both the global and the local features found in the mesoscopic optical images, the input size for the convolutional neural network is set to be 256 pixels by 256 pixels. All the input images are cropped randomly from the original images for training. The input for the denoising model during the prediction is cropped to the same size and arranged in a grid manner to recover a denoised whole slice. A subset 2000 images of the natural scene images bootstraps the process. The cropped images from the baseline and the target system are randomly divided for 10 fold cross-validation, among which 8 folds are for training, 1 fold is for validation, and 1 fold is for testing. There are 512 images from the baseline system (SI) and 512 images from the target system (TDI) set aside as the test set. The metrics reported below are collected over the test set. The image dataset is also augmented when used in training. The data augmentation methods include the horizontal flip and a x-y axis independent random zooming by 5%.

### C. RESULTS ON SYNTHETIC DATA

We first test our models over a set of synthetic noisy images. The backbone model used here is the DnUNet model illustrated in Fig. 2. The model is compared against other baseline algorithms for different noise type, as shown in Fig. 5

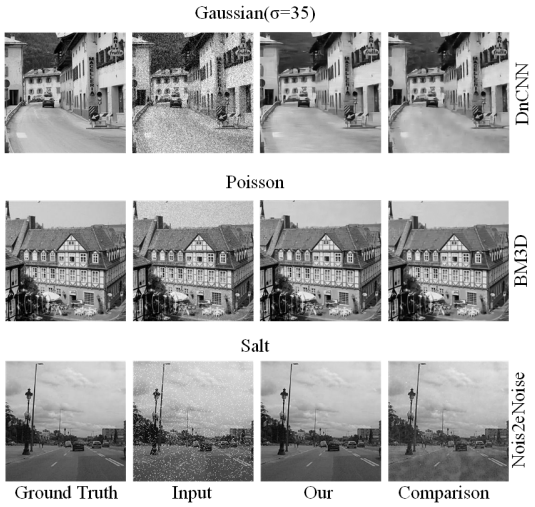



**FIGURE 5.** Some denoising examples of the DnUNet and the baseline algorithms on the synthetic noisy images.

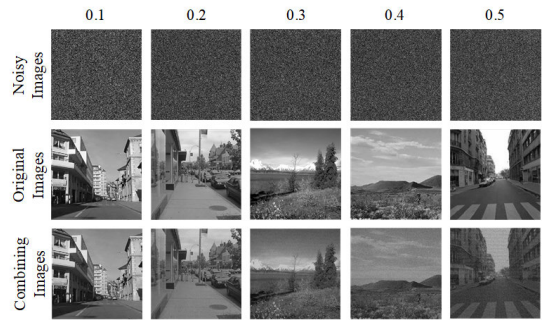



**FIGURE 6.** Noisy images from the natural scene dataset and the SI fixture area.

The natural images used in this experiment are the set of 2000 samples as described in Sec. III-B. Three example images are shown in Fig. 5. The first column is the clean image from the dataset; the second column is the synthetic noisy image following the Gaussian distributed noise, Poisson distributed noise, and salt-and-pepper noise; the proposed DnUNet results are in the third column and the baseline results for other three algorithms are shown in the last column. Over this synthetic experiment, we observe that the DnUNet has a better visual reconstruction of the details found in the ground truth. The drawback of the DnCNN and the noise2noise method is that their results have a block effect where the intensity is not uniform across the blocks, which can be observed in the road part of the last image on the first row and third row. BM3D, as a regional filter based method, it is harder for it to use the prior knowledge to recover the details using only the local information.

### D. BOOTSTRAP THE BASELINE DENOISING MODEL

The high-quality natural scene images are combined with the noisy images from the SI fixture areas described in Equ. 1 to form the training dataset. The result of the combination with different $\alpha_b$ is shown in Fig. 6. The images at the

**TABLE 1.** The PSNR of the 10 natural images with noisy fixture area overlay.

| | 0.1 | | | | | 0.3 | | | | | 0.5 | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| # | BM3D | N2N | DnCNN | DnUNet | UGAN | BM3D | N2N | DnCNN | DnUNet | UGAN | BM3D | N2N | DnCNN | DnUNet | UGAN |
| 1 | 29.36 | 29.80 | 36.35 | 40.95 | 41.50 | 20.30 | 19.99 | 30.82 | 31.43 | 35.30 | 15.72 | 16.44 | 23.02 | 20.89 | 25.90 |
| 2 | 28.70 | 29.77 | 35.95 | 39.96 | 40.62 | 20.02 | 19.84 | 29.89 | 30.05 | 31.66 | 15.69 | 16.19 | 23.26 | 21.57 | 25.01 |
| 3 | 28.45 | 29.58 | 36.90 | 39.96 | 0.99 | 20.15 | 19.98 | 29.30 | 29.61 | 31.26 | 15.78 | 16.39 | 22.55 | 20.65 | 24.06 |
| 4 | 27.63 | 28.07 | 36.72 | 41.12 | 41.55 | 18.39 | 18.23 | 30.20 | 31.73 | 33.97 | 14.01 | 14.44 | 23.26 | 22.00 | 25.41 |
| 5 | 32.60 | 33.43 | 37.18 | 42.28 | 43.10 | 23.98 | 23.38 | 31.36 | 30.21 | 36.25 | 19.57 | 20.66 | 22.98 | 21.77 | 26.91 |
| 6 | 30.43 | 31.21 | 37.24 | 41.30 | 41.38 | 21.58 | 21.31 | 31.11 | 30.81 | 33.41 | 17.30 | 17.71 | 22.39 | 21.87 | 23.25 |
| 7 | 29.19 | 30.40 | 36.40 | 40.12 | 40.94 | 20.92 | 20.73 | 28.74 | 29.60 | 31.72 | 16.49 | 16.98 | 22.70 | 21.00 | 22.96 |
| 8 | 31.23 | 33.55 | 36.66 | 40.43 | 41.57 | 24.10 | 23.98 | 29.78 | 29.73 | 31.68 | 19.91 | 20.69 | 22.86 | 21.92 | 24.49 |
| 9 | 28.73 | 29.76 | 35.14 | 40.35 | 41.02 | 20.23 | 20.08 | 28.48 | 29.79 | 31.80 | 15.76 | 16.24 | 23.00 | 21.53 | 24.35 |
| 10 | 27.35 | 28.15 | 34.10 | 39.46 | 41.00 | 18.57 | 18.43 | 28.25 | 30.80 | 32.09 | 14.19 | 14.60 | 23.64 | 22.09 | 24.29 |

**TABLE 2.** The SSIM of the 10 natural images with noisy fixture area overlay.

| | 0.1 | | | | | 0.3 | | | | | 0.5 | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| # | BM3D | N2N | DnCNN | DnUNet | UGAN | BM3D | N2N | DnCNN | DnUNet | UGAN | BM3D | N2N | DnCNN | DnUNet | UGAN |
| 1 | 0.975 | 0.989 | 0.985 | 0.989 | 0.986 | 0.955 | 0.942 | 0.964 | 0.967 | 0.967 | 0.913 | 0.920 | 0.951 | 0.939 | 0.947 |
| 2 | 0.954 | 0.987 | 0.981 | 0.986 | 0.987 | 0.921 | 0.913 | 0.929 | 0.941 | 0.938 | 0.866 | 0.863 | 0.884 | 0.878 | 0.844 |
| 3 | 0.949 | 0.988 | 0.982 | 0.937 | 0.987 | 0.910 | 0.903 | 0.923 | 0.938 | 0.928 | 0.848 | 0.852 | 0.879 | 0.867 | 0.867 |
| 4 | 0.971 | 0.989 | 0.985 | 0.989 | 0.988 | 0.946 | 0.937 | 0.950 | 0.962 | 0.961 | 0.899 | 0.899 | 0.924 | 0.926 | 0.873 |
| 5 | 0.982 | 0.992 | 0.990 | 0.993 | 0.991 | 0.966 | 0.961 | 0.971 | 0.979 | 0.978 | 0.935 | 0.940 | 0.959 | 0.946 | 0.959 |
| 6 | 0.960 | 0.978 | 0.982 | 0.990 | 0.988 | 0.909 | 0.900 | 0.943 | 0.950 | 0.956 | 0.846 | 0.840 | 0.867 | 0.880 | 0.833 |
| 7 | 0.938 | 0.973 | 0.980 | 0.985 | 0.986 | 0.888 | 0.881 | 0.913 | 0.923 | 0.927 | 0.824 | 0.820 | 0.851 | 0.852 | 0.826 |
| 8 | 0.947 | 0.987 | 0.982 | 0.988 | 0.989 | 0.912 | 0.911 | 0.923 | 0.938 | 0.932 | 0.862 | 0.862 | 0.873 | 0.864 | 0.847 |
| 9 | 0.958 | 0.987 | 0.982 | 0.989 | 0.989 | 0.916 | 0.910 | 0.931 | 0.949 | 0.946 | 0.842 | 0.836 | 0.881 | 0.879 | 0.876 |
| 10 | 0.955 | 0.986 | 0.980 | 0.988 | 0.989 | 0.910 | 0.904 | 0.934 | 0.947 | 0.945 | 0.833 | 0.826 | 0.877 | 0.884 | 0.856 |

first row show the noisy image from the fixture area. The images at the second row show the original images from the natural scene dataset. Images at the bottom row are the result of the combination of images from the first two rows. The figure shows that the noisy images have a regional-specific noise and a global noise pattern. This often arises when the whole brain slice is composed by overlapping tiles or the imaging system uses some non-uniform lighting and imaging for optical super-resolution, like the SI system. Another observation is that the intensity of the noisy images from the fixture area is not uniform and the learned denoising model should try to normalize it.

The denoising model performance for the SI system trained on the noisy images generated from the previous step is shown in Fig. 7. The value of the noise image weight $\alpha_b$ is shown along the x-axis, the first plot shows the PSNR score, and the second plot shows the SSIM score. The two scores are calculated for the baseline algorithms, including BM3D, DnCNN, and noise2noise. The vertical bar at each data point shows the relative scale for the score variance over the test dataset. Both the scores decrease as the noise strength increases as expected. From the result, we also can see that there is a performance gap between unsupervised algorithms and supervised algorithms. The performance gap is about 10dB for the PSNR score. The unsupervised denoising algorithms often seek the exception of the training images given a specific distribution assumption, while the supervised denoising ones strive to build up the mapping between the clean images and the noisy images. Among the supervised denoising models, our proposed DnUNet and GAN based algorithms also works well.

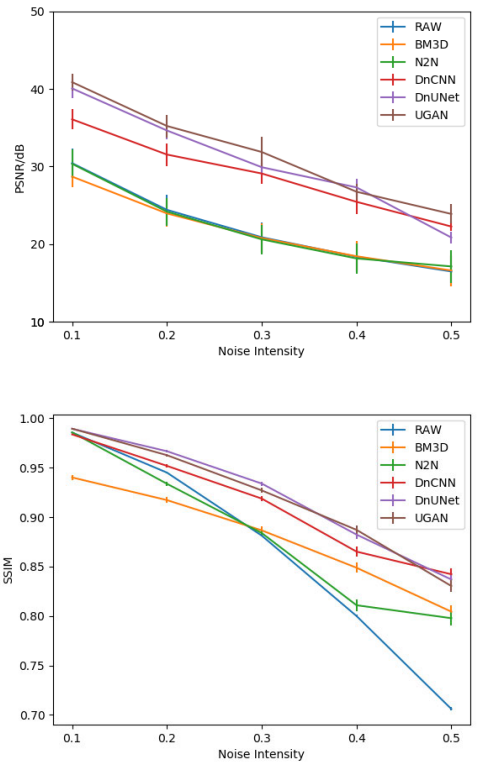



**FIGURE 7.** Denoising model performance for the SI system.

We also list the individual PSNR and SSIM score for the first 10 images in Tbl. 1 and Tbl. 2. The tables present the score under a different noise strength for each proposed and baseline algorithms. The first row of the header is the noise

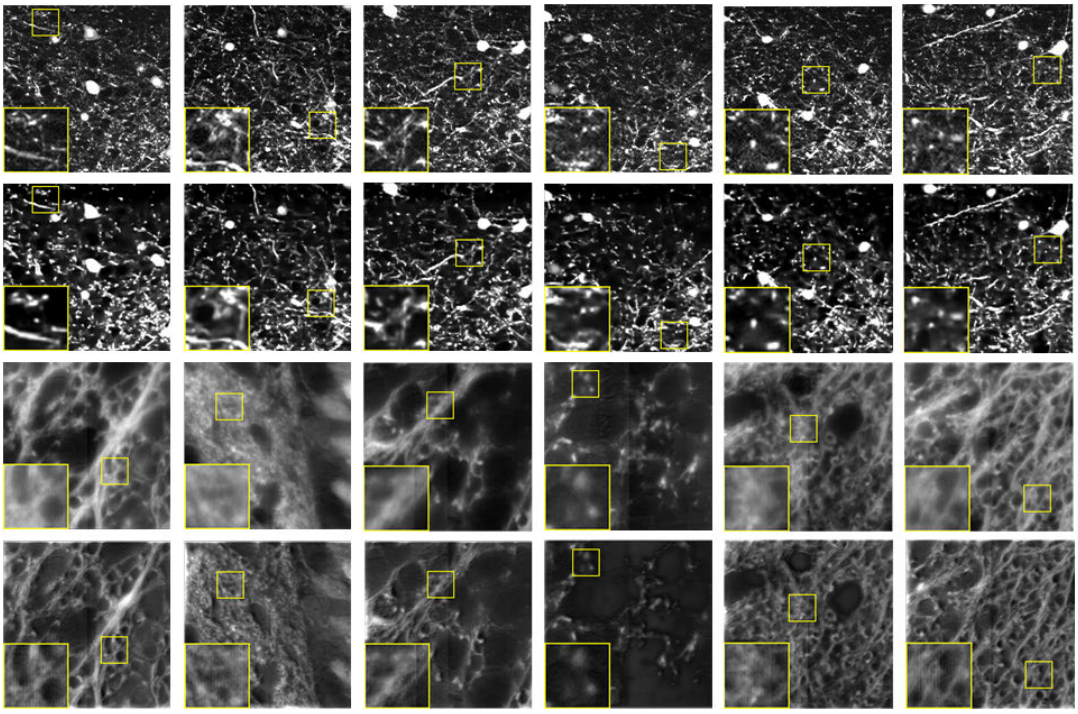


**FIGURE 8.** Image result at each phase of the CycleGAN architecture.

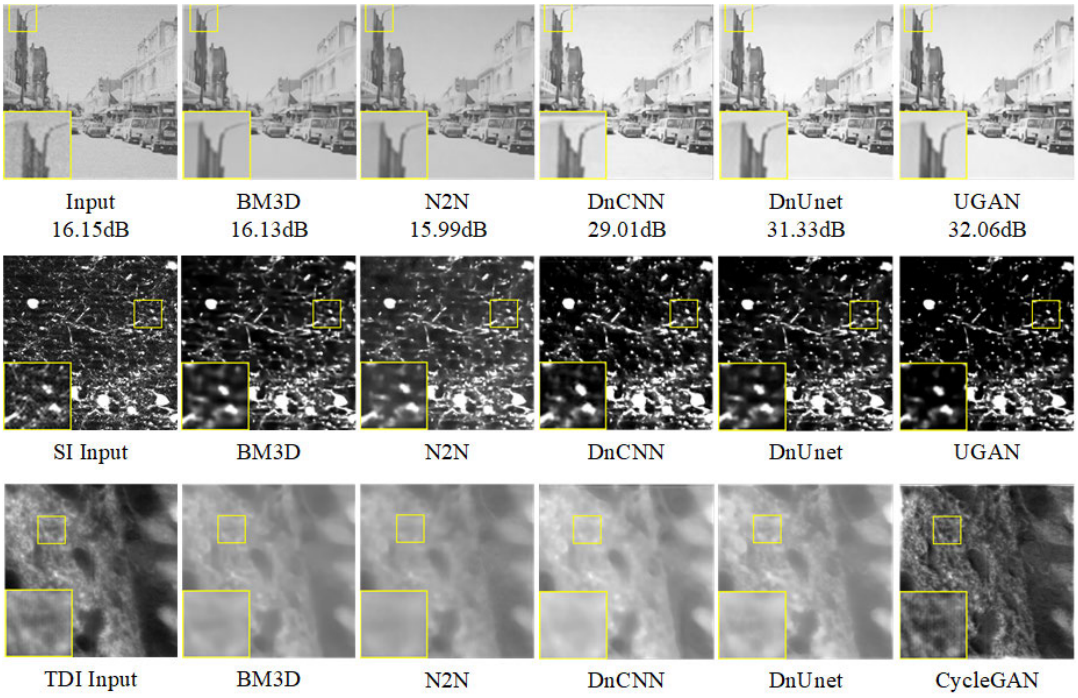



**FIGURE 9.** Denoising performance of SI denoising model on the TDI images.

strength and the second row of the header lists the algorithms. The list complements the information obtained from Fig. 7, that the performance gap between the supervised methods and the unsupervised methods is there not only over the dataset, but it is also true for each image. We also see UGAN's performance is relatively better than the proposed DnUNet's. This may suggest that more constraints from the CycleGAN helps the model to find the true clean image.

### E. DENOISING ACROSS MODALITIES WITH CYCLEGAN

We then present the denoising result during each phase of the CycleGAN architecture in Fig. 8. Images at the first row show the denoised image from the SI system, which is the system with a trained denoising model. The second row shows the corresponding generated TDI noise images of the first row with the help of the fixture area from TDI images, which is used to train the denoising model for the TDI system. The third row shows some original images from the TDI system. The fourth row shows the denoised image corresponding to the third row with the denoising model for TDI. Each larger yellow box shows the details of the smaller yellow box in the images. Although the PSNR and SSIM scores for this case are not available due to the fact that there is no ground truth clean images from the TDI system, we can still see the improved image quality of the denoised TDI images on the bottom row.

We also want to present the performance of the denoising model trained for the SI system on the TDI data in Fig. 9. This should demonstrate that, in the end, it is necessary to learn a new denoising model for a new imaging system. The images on the first row show the result of some competitive methods on the noisy natural scene images with their PSNR score. Images on the second row show the denoised result for the SI system through each method, which are trained for the SI system. Images on the last row show the same denoising model applied to the TDI system, except the CycleGAN approach. The result suggests that the denoising model for one imaging modality may not generalize well on a new imaging system. Thus we can choose to use CycleGAN alike architecture to transfer the denoising model to the new imaging modality

## IV. CONCLUSION

In this work, we proposed a CycleGAN based model to achieve the image denoising for the optical imaging system without the ground truth label across multiple imaging modalities. It takes advantage of the signal area and the fixture area of the whole brain slice to bootstrap a denoising model for one of the imaging modality. Then we build a GAN loop to convert the clean images in one modality to some noisy images in the other modality and denoise the converted noisy images to produce the clean images. This allows the strong prior knowledge from the other modality to guide the denoising model training procedure in the current modality.

The experimental results suggest that the proposed model outperforms several cutting edge methods for the denoising task. It demonstrates the possibility to train a denoising model for a new imaging system, when the clean-noisy image pair is not available for the new imaging system. The work also demonstrates the brain tissue consistency underlying multiple imaging modalities can help the optical neuroimaging analysis pipeline to migrate to a new system quickly without waiting for the human clean image labeling.

The work presented in the paper is also related with the domain adaption, where the vision model trained in one domain can be tuned to fit the similar purpose for the images from another domain. This is the case for our application where the denoising model trained for the original SI system needs to adapt to the new TDI system. However, the neuroscientists not only needs a new denoising model for the new imaging system, they would also be glad if there is a model can convert the new images back to the existing domain in order to either validate the new imaging system or use the existing downstream processing pipeline. The CycleGAN approach, usually recognized as an image translation task, can also convert the images between domains in a symmetric style for the neuroscientists.

One unexplored cost in building the CycleGAN based denoising models is the model efficiency. The approach introduced in this work has more parameters and more multiplication operations than the baseline algorithms. The denoising procedure usually sits ahead of the following optical neuroimaging data processing pipeline and a small model with a limited prediction timeframe would benefit the downstream jobs. The other future work that may improve the denoising performance further is to embed the denoising model into the whole optical neuroimaging data processing pipeline. As the pipeline is usually governed by the expert labelling dataset, the denoising model embedded in the whole pipeline will be guided by the labels in addition to the GAN loop described in this work and may improve its own performance because of this extra information.

## ACKNOWLEDGMENT

The authors would like to thank Xianbao Liu for providing the dataset and Xiaoquan Yang for suggestion. *(Tianfang Zhu and Yutong Han contributed equally to this work.)*

## REFERENCES

[1] S. Reardon, "A giant neuron found wrapped around entire mouse brain," *Nature News*, vol. 543, no. 7643, p. 14, 2017.

[2] A. Li, H. Gong, B. Zhang, Q. Wang, C. Yan, J. Wu, Q. Liu, S. Zeng, and Q. Luo, "Micro-optical sectioning tomography to obtain a high-resolution atlas of the mouse brain," *Science*, vol. 330, no. 6009, pp. 1404–1408, Dec. 2010.

[3] X. Qi, T. Yang, L. Li, J. Wang, S. Zeng, and X. Lv, "Fluorescence micro-optical sectioning tomography using acousto-optical deflector-based confocal scheme," *Neurophotonics*, vol. 2, no. 4, Oct. 2015, Art. no. 041406.

[4] S. W. Hell, S. J. Sahl, M. Bates, X. Zhuang, R. Heintzmann, M. J. Booth, J. Bewersdorf, G. Shtengel, H. Hess, P. Tinnefeld, and A. Honigmann, "The 2015 super-resolution microscopy roadmap," *J. Phys. D, Appl. Phys.*, vol. 48, no. 44, 2015, Art. no. 443001.

[5] G. Lepage, J. Bogaerts, and G. Meynants, "Time-delay-integration architectures in CMOS image sensors," *IEEE Trans. Electron Devices*, vol. 56, no. 11, pp. 2524–2533, Nov. 2009.

[6] Q. Sun, X. Li, M. Ren, M. Zhao, Q. Zhong, Y. Ren, P. Luo, H. Ni, X. Zhang, C. Zhang, J. Yuan, A. Li, M. Luo, H. Gong, and Q. Luo, "A whole-brain map of long-range inputs to GABAergic interneurons in the mouse medial prefrontal cortex," *Nature Neurosci.*, vol. 22, no. 8, pp. 1357–1370, Aug. 2019.

[7] T. Quan, H. Zhou, J. Li, S. Li, A. Li, Y. Li, X. Lv, Q. Luo, H. Gong, and S. Zeng, "Neurogps-tree: Automatic reconstruction of large-scale neuronal populations with dense neurites," *Nature methods*, vol. 13, no. 1, p. 51, 2016.

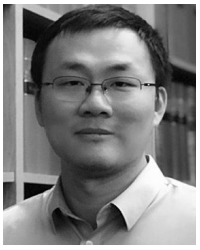
[8] Q. Luo, "Visible brain-wide networks at single-neuron resolution with micro-optical sectioning tomography," in *CLEO: Applications and Technology*. Washington, DC, USA: Optical Society of America, 2014, Art. no. AF1B-5.

[9] H. C. Burger, C. J. Schuler, and S. Harmeling, "Image denoising: Can plain neural networks compete with BM3D?" in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2012, pp. 2392–2399.

[10] A. M. Wink and J. B. J. M. Roerdink, "Denoising functional MR images: A comparison of wavelet denoising and Gaussian smoothing," *IEEE Trans. Med. Imag.*, vol. 23, no. 3, pp. 374–387, Mar. 2004.
[11] T.-E. Kam, X. Wen, B. Jin, Z. Jiao, L.-M. Hsu, Z. Zhou, Y. Liu, K. Yamashita, S. C. Hung, W. Lin, and H. Zhang, "A deep learning framework for noise component detection from resting-state functional MRI," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Cham, Switzerland: Springer, 2019, pp. 754–762.
[12] S. Laine, T. Karras, J. Lehtinen, and T. Aila, "High-quality self-supervised deep image denoising," in *Proc. 33rd Conf. Neural Inf. Process. Syst. (NeurIPS)*, 2019, pp. 6970–6980.
[13] A. Krull, T.-O. Buchholz, and F. Jug, "Noise2void-learning denoising from single noisy images," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 2129–2137.
[14] J. Batson and L. Royer, "Noise2Self: Blind denoising by self-supervision," 2019, *arXiv:1901.11365*. [Online]. Available: http://arxiv.org/abs/1901.11365
[15] C. Tomasi and R. Manduchi, "Bilateral filtering for gray and color images," in *Proc. 6th Int. Conf. Comput. Vis.*, 1998, pp. 839–846.
[16] L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Phys. D, Nonlinear Phenomena*, vol. 60, nos. 1–4, pp. 259–268, Nov. 1992.
[17] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising by sparse 3-D transform-domain collaborative filtering," *IEEE Trans. Image Process.*, vol. 16, no. 8, pp. 2080–2095, Aug. 2007.
[18] J. Mairal, F. Bach, J. Ponce, G. Sapiro, and A. Zisserman, "Non-local sparse models for image restoration," in *Proc. IEEE 12th Int. Conf. Comput. Vis.*, Sep. 2009, pp. 2272–2279.
[19] J. Xie, L. Xu, and E. Chen, "Image denoising and inpainting with deep neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2012, pp. 341–349.
[20] X.-J. Mao, C. Shen, and Y.-B. Yang, "Image restoration using very deep convolutional encoder-decoder networks with symmetric skip connections," in *Proc. 30th Int. Conf. Neural Inf. Process. Syst. (NIPS)*, 2016, pp. 2810–2818.
[21] K. Zhang, W. Zuo, Y. Chen, D. Meng, and L. Zhang, "Beyond a Gaussian denoiser: Residual learning of deep CNN for image denoising," *IEEE Trans. Image Process.*, vol. 26, no. 7, pp. 3142–3155, Jul. 2017.
[22] Y. Chang, L. Yan, M. Chen, H. Fang, and S. Zhong, "Two-stage convolutional neural network for medical noise removal via image decomposition," *IEEE Trans. Instrum. Meas.*, vol. 69, no. 6, pp. 2707–2721, Jun. 2020.
[23] Y. Luo, J. Zou, C. Yao, X. Zhao, T. Li, and G. Bai, "HSI-CNN: A novel convolution neural network for hyperspectral image," in *Proc. Int. Conf. Audio, Lang. Image Process. (ICALIP)*, Jul. 2018, pp. 464–469.
[24] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *Proc. Adv. Neural Inf. Process. Syst.*, 2014, pp. 2672–2680.
[25] M. Mirza and S. Osindero, "Conditional generative adversarial nets," 2014, *arXiv:1411.1784*. [Online]. Available: http://arxiv.org/abs/1411.1784
[26] M. Arjovsky, S. Chintala, and L. Bottou, "Wasserstein GAN," Tech. Rep., 2017.
[27] K. J. Halupka, B. J. Antony, M. H. Lee, K. A. Lucy, R. S. Rai, H. Ishikawa, G. Wollstein, J. S. Schuman, and R. Garnavi, "Retinal optical coherence tomography image enhancement via deep learning," *Biomed. Opt. Exp.*, vol. 9, no. 12, pp. 6205–6221, Dec. 2018.
[28] Y. Huang, Z. Lu, Z. Shao, M. Ran, J. Zhou, L. Fang, and Y. Zhang, "Simultaneous denoising and super-resolution of optical coherence tomography images based on generative adversarial network," *Opt. Exp.*, vol. 27, no. 9, pp. 12289–12307, 2019.
[29] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 2242–2251.
[30] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Cham, Switzerland: Springer, 2015, pp. 234–241.
[31] A. Oliva and A. Torralba, "Modeling the shape of the scene: A holistic representation of the spatial envelope," *Int. J. Comput. Vis.*, vol. 42, no. 3, pp. 145–175, 2001.
[32] J. Lehtinen, J. Munkberg, J. Hasselgren, S. Laine, T. Karras, M. Aittala, and T. Aila, "Noise2Noise: Learning image restoration without clean data," 2018, *arXiv:1803.04189*. [Online]. Available: http://arxiv.org/abs/1803.04189
[33] H. Gong, D. Xu, J. Yuan, X. Li, C. Guo, J. Peng, Y. Li, L. A. Schwarz, A. Li, B. Hu, B. Xiong, Q. Sun, Y. Zhang, J. Liu, Q. Zhong, T. Xu, S. Zeng, and Q. Luo, "High-throughput dual-colour precision imaging for brain-wide connectome with cytoarchitectonic landmarks at the cellular level," *Nature Commun.*, vol. 7, no. 1, p. 12142, Nov. 2016.

**TIANFANG ZHU** received the B.E. degree from Hubei University, Wuhan, China, in 2020. He is currently pursuing the master's degree with the Huazhong University of Science and Technology, Wuhan, China. His main research interests include biomedical image analysis and machine learning.

**YUTONG HAN** received the B.E. degree from the Wuhan Institute of Technology, Wuhan, China, in 2015. She is currently pursuing the Ph.D. degree with the Huazhong University of Science and Technology, Wuhan, China. Her main research interest includes optical neuroimaging.



**ANAN LI** received the B.E. and Ph.D. degrees in biomedical engineering from the Huazhong University of Science and Technology, in 2005 and 2011, respectively. In 2011, he joined the Wuhan National Laboratory for Optoelectronics. His research interests include processing, analysis, and visualization of the TB-sized neural images.

**JING YUAN** received the Ph.D. degree in electrical engineering from the Huazhong University of Science and Technology, in 2007. In 2010, she joined the Wuhan National Laboratory for Optoelectronics. Her research interest includes developing optical system for the whole brain neuroimaging.

**GONG HUI** received the B.E. degree in engineering from Xidian University, in 1986. She joined the Huazhong University of Science and Technology, in 1986, and the Wuhan National Laboratory for Optoelectronics, in 2005. Her research interest includes mesoscopic scale whole brain connectome.

**YUE GUAN** (Member, IEEE) received the Ph.D. degree in electrical engineering from Northeastern University, Boston. He is currently a Teacher with the Huazhong University of Science and Technology, Wuhan, China. His research interests include machine learning, pattern recognition, and neuroscience inspired artificial intelligence.