High-throughput, high-resolution deep learning microscopy based on registration-free generative adversarial network

HAO ZHANG,[1,6] CHUNYU FANG,[1,6] XINLIN XIE,[1] YICONG YANG,[1] WEI MEI,[3] DI JIN,[4,7] AND PENG FEI[1,2,5,*]

[1]*School of Optical and Electronic Information, Huazhong University of Science and Technology, Wuhan, 430074, China*
[2]*Britton Chance Center for Biomedical Photonics, Wuhan National Laboratory for Optoelectronics, Huazhong University of Science and Technology, Wuhan, 430074, China*
[3]*Department of Anesthesiology, Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, 430030, China*
[4]*Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge, MA, 02139, USA*
[5]*Shenzhen Huazhong University of Science and Technology Research Institute, Shenzhen, 518000, China*
[6]*equally contributing author*
[7]*jindi15@mit.edu*
*feipeng@hust.edu.cn

Abstract: We combine a generative adversarial network (GAN) with light microscopy to achieve deep learning super-resolution under a large field of view (FOV). By appropriately adopting prior microscopy data in an adversarial training, the neural network can recover a high-resolution, accurate image of new specimen from its single low-resolution measurement. Its capacity has been broadly demonstrated via imaging various types of samples, such as USAF resolution target, human pathological slides, fluorescence-labelled fibroblast cells, and deep tissues in transgenic mouse brain, by both wide-field and light-sheet microscopes. The gigapixel, multi-color reconstruction of these samples verifies a successful GAN-based single image super-resolution procedure. We also propose an image degrading model to generate low resolution images for training, making our approach free from the complex image registration during training data set preparation. After a well-trained network has been created, this deep learning-based imaging approach is capable of recovering a large FOV (~95 mm^2) enhanced resolution of ~1.7 μm at high speed (within 1 second), while not necessarily introducing any changes to the setup of existing microscopes.

1. Introduction

The imaging throughput of a conventional optical microscope is typically limited to megapixels, regardless of the magnification and numerical aperture used [1,2]. As a result, compromise often exists between achieving a high resolution and maintaining a large field-of-view (FOV). However, nowadays high-resolution mapping of entire large specimens is increasingly desirable for life science applications such as tissue pathology, hematology, digital histology and neuron science [3,4]. In order to precisely interpret cellular events throughout entire samples, global structures and local details spanning from micro- to meso-scale need to be continuously measured and quantitatively analyzed at the same time [5]. Development of sophisticated mechanical scanning microscope is a commonly-used way to address this challenge, artificially increasing the throughput of the microscope by stitching multiple high-resolution tiles into a panoramic image [6]. Besides this mechanical approach that requires precise control over actuation and optical alignment, recent super resolution

(SR) techniques present a computational way to increase the space-bandwidth product of a microscope platform [1,7–19] For instance, pixel super resolution (PSR) represents a class of spatial domain techniques that can fuse multiple large FOV, low resolution measurements with sub-pixel shifts into a high resolution image [17,18]. On the other hand, several frequency domain methods, e.g., Fourier ptychographic microscopy (FPM) [1], synthetic aperture microscopy [7–10] and structured-illumination microscopy [20,21], produce a resolution-enhanced image by stitching together a number of variably illuminated, low-resolution images in Fourier domain. Despite offering unique imaging capabilities with scalable SBP, these methods, however, all require special hardware setup and complex computation on multiple frames. Nevertheless, another type of technique, named single image super resolution (SISR), has been widely applied in microscopy without these constraints. It aims at the reconstruction of a high-resolution (HR) images with rich details from single low-resolution (LR) image. For this technique, the conventional widely used method is the example-based approach [22,23], which works by replacing the LR information with the HR patches searched out in the example dictionary. Although SISR requires neither high-resolution imaging hardware architecture nor intensive computation resource, the quality of reconstructed images remains suboptimal as compared to the multi-frame methods. The recent advent of deep learning neural network is providing another way to realize more effective SISR. Apart from its success in medical diagnosis like carcinoma detection, gliomas grading, histopathological segmenting and classifying [24–26], deep learning has been used in the super-resolution in bright-field microscopy [27,28] as well as fluorescence microscopy [29–32]. The most recent model that utilizes the generative adversarial network (GAN) for better visual details enhancement, has reached remarkable resolution enhancement [29,32]. However these methods require an extra image registration between high-resolution and low-resolution training pairs captured under different magnifications. Considering a pixel-wise error function is the most common practice in super resolution, the accuracy of registration could affect the performance of the neural network.

Here we present a deep learning-based super resolution approach that is free from registration during training process, meanwhile capable of providing significant resolution enhancement for conventional microscopy, without the need of acquiring a plurality of frames or retrofitting existing optical systems [33]. This imaging method uses data sets that consist of high-resolution measurements and their low-resolution simulations to train a GAN model. We carefully model the image degradation of the microscope system to generate low-resolution trial images from measured high-resolution source images, thereby eliminating the need of complicated alignment between the high- and low-resolution pairs. As long as the network training is accomplished, the network is capable of using single low-resolution measurement of a new specimen to recover its high-resolution, large FOV image. We demonstrate the efficiency of this registration-free GAN microscopy (RFGANM) method with bright-field image of USAF resolution target, color image of whole pathological slides, dual-channel fluorescence image of fibroblast cells, and light-sheet fluorescence image of a whole mouse brain, verifying that it's widely applicable to various microscopy data. By taking a few example images as the references and applying a GAN deep-learning procedure, we can transform a conventional optical microscope into a high-resolution (~1.7 μm), wide-FOV (~95 mm^2) microscope with a final effective SBP of 0.13 gigapixels. Furthermore, unlike the training stage that must be performed on GPUs to greatly reduce the time cost, reconstructing procedure can work readily with an ordinary CPU device in still acceptable time of several minutes per image. This underlying advantage renders RFGANM a robust platform that allows multiple applications to be followed once after a well-trained SR artificial intelligence based system is established. In the following, we will briefly describe the RFGANM operation and experimental set-up, discuss how to perform the network training and inference process, discuss its imaging applications in a variety of biomedical samples, and demonstrate

how RFGANM can benefit bio-medical analysis such as cancer diagnosis, cell counting in pathological section images and neuron profiling in light sheet image of mouse brain.

2. Results

2.1 Deep learning based image super resolution reconstruction

A classic GAN model [34] that consists of a generator and a discriminator, is used to "learn" the various types of microscopy data from scratch. Figure 1 illustrates the network training and inference process. We establish its capability of mapping from a LR image to a HR reconstruction as shown in Fig. 1(a). Firstly multiple HR images of the example specimen are captured under high-magnification objective (Fig. 1(a), step 1), then through accurately modeling the transfer function of the microscope system, we can obtain the down-sampled, blurred images of the example specimen directly via simulation (Fig. 1(a), step 2). Based on its currently-learned parameters, the generator creates resolution-enhanced reconstructions of LR simulations in each training iteration (Fig. 1(a), step 3). The differences between the generator outputs and the realistic HR images are calculated using the mean squared error (MSE), denoted as the content loss function of the generator (Fig. 1(a), step 4). Besides the generator, GAN includes an additional discriminator that aims to evaluate the reliability of the generator. This discriminator makes a judgement on whether an image is a reconstruction by the generator or a realistic high-resolution measurement, after they are randomly input (Fig. 1(a), step 5). An adversarial loss is created to estimate the accuracy of the discriminator's judgement. It iteratively optimizes the discriminator, aiming at an enhanced capability on making correct decision. Also, the adversarial loss together with the content loss are used to optimize the generator, pushing it towards the direction that generates more perceptually realistic outputs which can further fool the discriminator. This adversarial training process thereby promotes the accuracy of both the generator and the discriminator. The training process can be terminated when the generator produces results that the discriminator can hardly tell from the realistic HR images. Then in the inference phase, a LR measurement of sample, which is excluded from the training data set, is divided into several patches and fed into the well-trained generator (Fig. 1(b), step 6). The generator is capable of recovering high frequency information for each patch, based on the prior GAN training. These quality-improved patches are finally stitched into one gigapixel image of the sample that encompasses high-resolution details as well as large FOV (Fig. 1(b), step 7). The aforementioned image reconstruction process is illustrated in Fig. 1(b), and the overall implementation process of our approach is illustrated in Fig. 5(a). It is noteworthy that usually the GAN training is required only once, and then applicable to the recovery of multiple samples with similar type of signals.
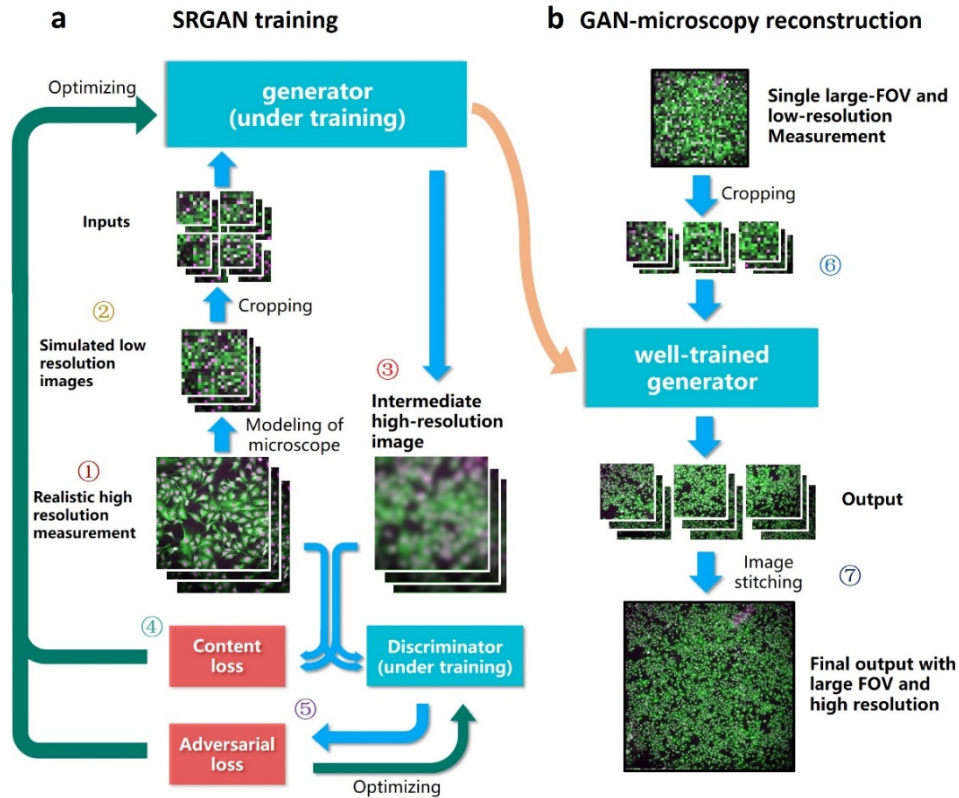
Fig. 1. Principle of RFGANM procedure. A modified GAN is used to iteratively learn the microscopy data till the goal of high-quality output is reached. (a) The training process of the RFGAN. (b) The RFGANM reconstruction procedure, using a large-FOV, low-resolution measurement of new sample.

2.2 Image degrading model

It is widely accepted that the performance of a neural network relies heavily on the training data set, where there are LR images as inputs and HR images as targets for super resolution task. These LR and HR image pairs for training can be obtained in two ways. Most intuitively, both LR and HR images are experimentally captured with microscope. However, since the LR and HR image pairs are taken under different magnifications, image cutting and registration techniques must be used to match the FOV and remove the unavoidable distortion. Therefore, the performance of image registration is the key to the quality of training data, which is mainly based on feature detection and matching. Unfortunately, in cases of cell, tissue and brain imaging, a great deal of feature details is lost in LR images compared with the corresponding HR images due to the down-sampling process, leading to a high failure rate of image registration. Even though we have used a decent and standard image registration procedure, the mismatch between LR and HR images happens a lot, which significantly deteriorates the quality of training data set.

Instead of capturing LR and HR images under different magnifications and then aligning them, we can apply an image degrading model to the captured HR images to generate the simulated LR images. In a nutshell, the LR images for training are directly down-sampled from the HR images, so we can guarantee that the two images share the same FOV. To make sure that our model trained on the simulated LR images can still well super-resolve the experimentally captured LR images, the image degrading model we used should be able to

produce a simulated LR image as close to the captured LR counterpart as possible. The degrading process of conventional microscopy system can be described as

$$I_m = D(K * I) + N \tag{1}$$

where I is the continuous real intensity distribution of the sample to be imaged; K is the point spread function of the optical system, represented as a Gaussian convolution kernel; operator $*$ is the convolution between I and K; D acting on the convolution results denotes the discretization by the camera sensor; N denotes the additive Gaussian white noise, mainly contributed by the statistic thermal noise of the CCD/CMOS sensor; and I_m is the digital measurement we obtained, which is the discrete, decimated approximation of I.

In practice, the real distribution I is approximated by a high-resolution measurement obtained under high-magnification optics. Discretization operation D is thereby a down-sampling on the digital image. There are two parameters to be optimized: size of the Gaussian kernel in the convolution step and variance of the noise distribution. This trail-and-error procedure is visualized in Fig. 2. Taking the degradation of a 20 × cell image as an example, we de-noise the 4x measurement, comparing it with the blurred, down-sampled 20 × measurement to figure out the proper sigma value of the Gaussian kernel. After an optimal sigma value is found, we fix it and add noise with different variance, comparing the degraded result with the original 4x measurement to identify the proper variance. After the degrading model being established, we verify it multiple times. The difference between the simulation and the measurement is directly computed via the pixel-wise subtraction, as shown in Fig. 2(g). Except for slight mismatch of the cell shape caused by the imaging aberration, the differences are nearly Gaussian white noise, demonstrating the successful application of the degrading model.
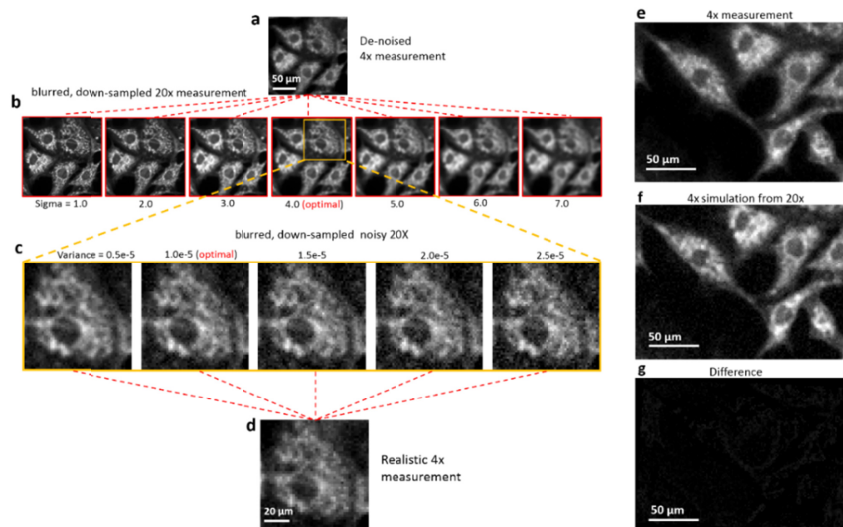

Fig. 2. Generation of simulated LR inputs via a degradation model. (a) Denoised 4 × LR measurement used for finding the optimized blurring kernel of microscope. (b) The blurred and down-sampled images from 20 × measurement, by different sizes of blurring kernel (Sigma value). (c) Blurred image with optimal blurring is processed with different additive noises and compared to the realistic 4 × measurement. The best matched level of noise is found as a result. (e-g) A 4 × measurement is subtracted by the optimized 4 × simulation to verify the efficacy of the model.

After generating the simulated LR images using the captured HR measurements, we cut the images into small pieces to compose the training data set, of which the LR pieces are 96 × 96 pixels and HR pieces are 384 × 384 pixels since smaller image size is more compatible to

the limited GPU memory. Each LR piece has a corresponding HR one with the same FOV. We also adopt several widely-used data augmentation methods, such as translation, rotation and flipping, to further expand the training data set.

Implementing appropriate GAN-training with simulations generated by an accurate degrading model is crucial for the high quality SR outputs. The accurate and not accurate LR simulations used as the training data result in very different network outputs, as shown in Fig. 3. Taking the same LR measurement as the input, network trained with finely-tuned simulations generates far better results than the one trained with inaccurate simulations.
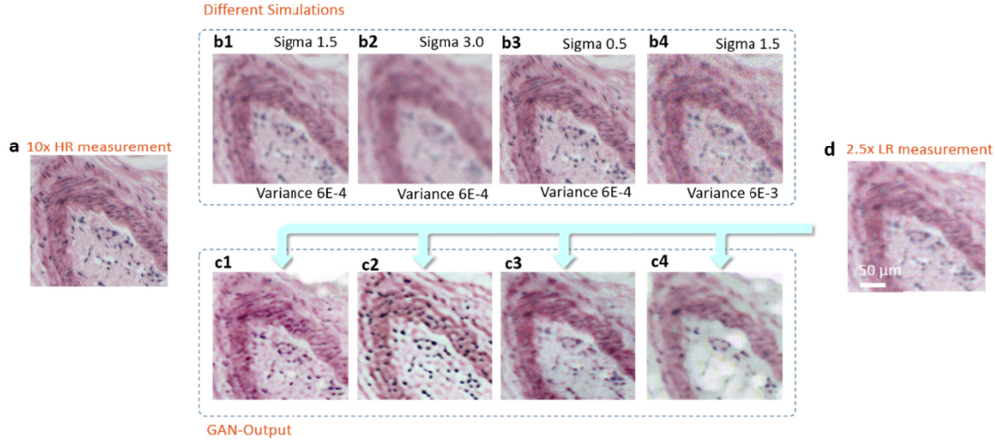

Fig. 3. Network outputs with different degrading parameters. (a) A 10 × measurement of a human stomach tissue slice. (b1-b4) LR simulations degraded from (a), with different parameters applied. Sigma is the standard deviation of the Gaussian blur kernel and variance denotes the variance of the white noise distribution. (c1-c4) After the network being trained by different simulation data, it correspondingly generates different reconstruction results that are recovered from the same LR measurement (d).

2.3 Neural net architecture and loss functions

The model architecture of the GAN is illustrated in Fig. 4. In the generator, the input image is first processed by one convolution layer of 3 x 3 kernel followed by ReLU activation and then goes through 16 residual blocks with identical layout. In each residual block, there are two convolutional layers with small 3 x 3 kernels and 64 feature maps followed by batch-normalization (BN) layers and ReLU as the activation function, and the input and output are element-wise added via shortcut connections. To enhance the residual learning, the output feature map of the first convolution layer and that of the final residual block is also element-wise added. We increase the resolution of image using two layers of sub-pixel convolution suggested by Shi et al. [35].

To optimize the generator, we designed the perceptual loss function that is the weighted sum of the MSE loss ℓ_{MSE}^{G}, the feature reconstruction loss ℓ_{feat}^{G} and the adversarial loss ℓ_{adv}^{G} as proposed by Ledig et al. [36]:

$$\ell^{G} = \frac{1}{N}\sum_{n=1}^{N}\left(\ell_{MSE}^{G}\left(G_{\theta_{G}}\left(\Delta\left(I_{n}^{HR} \right) \right), I_{n}^{HR} \right) + 20^{-6}\,\ell_{feat}^{G}\left(G_{\theta_{G}}\left(\Delta\left(I_{n}^{HR} \right) \right), I_{n}^{HR} \right) + 10^{-3}\,\ell_{adv}^{G}\left(G_{\theta_{G}}\left(\Delta\left(I_{n}^{HR} \right) \right) \right) \right)$$

$$(2)$$

where $G_{\theta_{G}}$ is the generator parameterized by θ_{G}, I^{HR} is the high resolution measurement and Δ is the degeneracy course by our degrading model. In this equation, the MSE loss measures the pixel-wise difference between the output super-resolved image and the target high-resolution image, calculated as:

$$\ell_{MSE}^{G}\left(G_{\theta_G}\left(\Delta\left(I_n^{HR}\right)\right), I_n^{HR}\right) = \frac{1}{r^2 WH}\sum_{x=1}^{rW}\sum_{y=1}^{rH}\left(I_{x,y}^{HR} - G_{\theta_G}\left(\Delta\left(I_n^{HR}\right)\right)_{x,y}\right)^2 \qquad (3)$$

Besides using the GAN framework to encourage perceptual similarity, we further used the special feature reconstruction loss function proposed by Johnson et al. [37]. Let $\phi_j(x) \in R^{W_j \times H_j \times C_j}$ be the activations of the jth convolution layer of the VGG19 network described in Simonyan and Zisserman [38] when processing the image. Then the feature reconstruction loss is defined by the Euclidean distance between the feature representations of the reconstructed image $G_{\theta_G}\left(\Delta\left(I_n^{HR}\right)\right)$ and the reference image I^{HR}:

$$\ell_{feat/j}^{G}\left(G_{\theta_G}\left(\Delta\left(I_n^{HR}\right)\right), I_n^{HR}\right) = \frac{1}{W_j H_j}\sum_{x=1}^{W_j}\sum_{y=1}^{H_j}\left(\phi_j\left(I^{HR}\right)_{x,y} - \phi_j\left(G_{\theta_G}\left(\Delta\left(I_n^{HR}\right)\right)\right)_{x,y}\right)^2 \quad (4)$$

where j in our experiments was set to 12.

In addition to the losses described so far, we also need to add the adversarial component of our GAN for the generative side to the perceptual loss. It is defined based on the probabilities of the discriminator over the reconstructed samples as:

$$\ell_{adv}^{G}\left(G_{\theta_G}\left(\Delta\left(I_n^{HR}\right)\right)\right) = -\log D_{\theta_D}\left(G_{\theta_G}\left(\Delta\left(I_n^{HR}\right)\right)\right) \qquad (5)$$

where D_{θ_D} is the discriminator parameterized by θ_D. For better gradient computation stability [34], we minimize $-\log D_{\theta_D}\left(G_{\theta_G}\left(\Delta\left(I_n^{HR}\right)\right)\right)$ instead of $-\log\left(1 - D_{\theta_D}\left(G_{\theta_G}\left(\Delta\left(I_n^{HR}\right)\right)\right)\right)$.

As for the discriminator, It first contains 8 convolutional layers with 4 x 4 kernels followed by BN layers and LeakyReLU (α = 0.2) activation (except that the first convolutional layer does not come with BN). Through these 8 layers, the feature map dimension first increases gradually by a factor of 2 from 64 to 2048, then decreases by the same factor to 512. Strided convolutions with stride of 2 are used to reduce the image resolution each time the number of features is doubled. Afterwards, the network is followed by a residual block that contains three convolutional layers followed by BN and LeakyReLU activation. Finally, the resulted 512 feature maps are flattened and connected by one dense layer and a sigmoid activation function to obtain the final probability over whether the input image is natural or not. The network is trained by minimizing the following loss function:

$$\ell^{D} = \frac{1}{N}\sum_{n=1}^{N}\left(-\log\left(D_{\theta_D}\left(I_n^{HR}\right)\right) - \log\left(1 - D_{\theta_D}\left(G_{\theta_G}\left(\Delta\left(I_n^{HR}\right)\right)\right)\right)\right) \qquad (6)$$
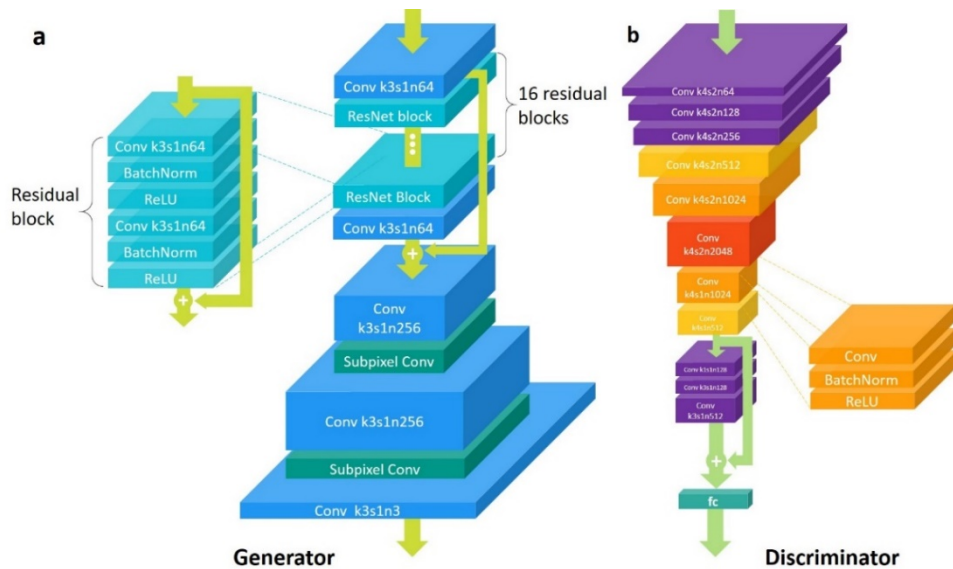
Fig. 4. The architecture of GAN. (a) The architecture of the generator. Conv and ResNet is the abbreviation of Convolutional layer and Residual network block. The parameters of the convolutional layers is given in the format "*k-s-n*", where *k* is the kernel size, *s* is the strides and *n* is the number of feature maps (i.e. the output channels of the layer). The depth of each convolutional layer roughly denotes the number of its feature maps, and the lateral dimensions denotes the size of its input. Totally, there are 16 residual blocks in the generator. (b) The architecture of the discriminator. Each convolutional layer in the discriminator is the combination of a convolution layer, a batch normalization operation and a ReLU activation function.

2.4 Characterization of super-resolution GAN

The capability of GAN is first characterized through imaging a negative USAF resolution target (Thorlabs R3L3S1N) with highest resolution of 228 line pairs per mm (lpm). We captured HR and corresponding LR images under a macro microscope (Olympus MVX10) with 10 × and 2 × total magnifications, respectively. Due to the simple pattern of the test target, an image registration was applied to match their corresponding FOV, forming strictly aligned HR and LR pairs for the GAN training. Considering the limited number of experimentally obtained samples, we applied a geometric transformation, such as translation and rotation, to these paired images to further expand the data set. Finally, 1008 groups of HR and LR pairs were imported into the GAN network for training. Another large FOV, LR measurement was used to validate the converged network (Fig. 5(b)). As shown in Fig. 5(c), (d), the 5x-enhanced reconstructions have a significant improvement compared to the raw images. Due to the small magnification factor as well as limited numerical aperture, the raw 2 × image can hardly discern the high-frequency stripes in USAF target (Fig. 5(b), cyan box for 114, and orange box for 228 lpm). The RFGANM reconstruction, in contrast, has resolved the finest part of USAF target (Fig. 5(d2), 228 lpm). The GAN-reconstruction results are further compared with the realistic measurement under a 10 × magnification (Figs. 5(c3) and 5(d3)), showing a good structural similarity (SSIM) to the high-resolution ground truth. The linecuts through the resolved line pairs (Fig. 5(d)) by each method are quantified in Fig. 5(e), revealing a substantially improved resolution by GAN.
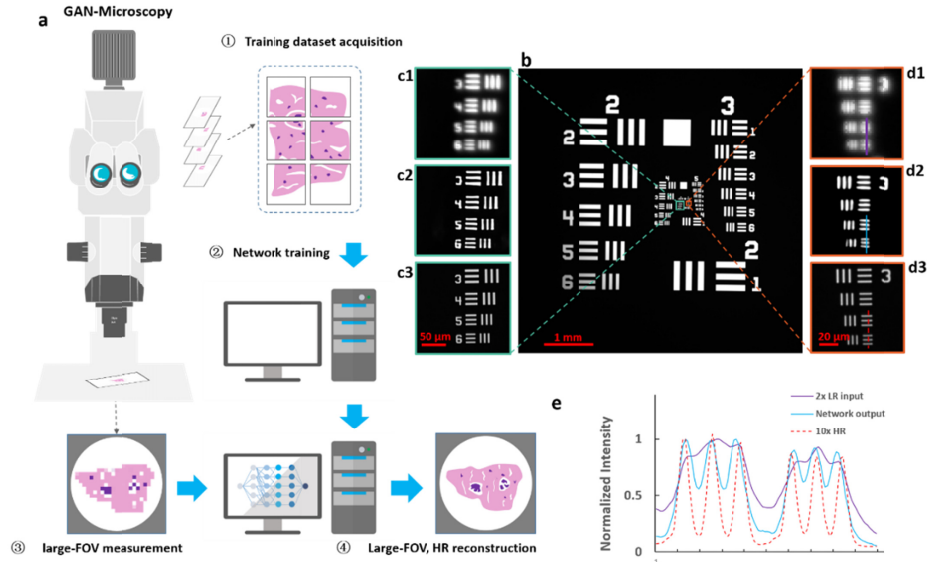
Fig. 5. Resolution characterization of GAN using USAF target. (a) The initial implementation of RFGANM, including HR training data acquisition, GAN network training, LR image acquisition of the sample, and HR reconstruction. (b) Image of the USAF resolution target taken under 2 × low-magnification of a macro-zoom microscope. (c1-d1) Magnified views of the raw image, with a pixel size of 2.13 μm. (c2-d2) Reconstructions of a well-trained GAN-generator, with an enhancement factor of 5 (reconstructed pixel size, 0.425 μm). (c3-d3) Corresponding high-resolution images taken under 10 × magnification. (e) Intensity plot of linecuts (shown in (d1-d3)) for each method, indicating that RFGANM prototype provides substantively improved contrast and resolution (FWHM ~1.7 μm) that potentially enable subcellular-level imaging across a centimeter large scale.

2.5 RFGANM of dual-channel fluorescent BPAE cells

We demonstrate our prototype by acquiring a wide-FOV dual-channel fluorescent image of a bovine pulmonary artery endothelial (BPAE) cell slide, as shown in Fig. 6(a). The DAPI and Alexa Fluor 488 tagged nucleus and micro-filaments of the cells were imaged, trained and validated separately and finally merged into a pseudo-color display. To circumvent the high-demanding alignment between HR- and LR- measurements for their complicated subcellular patterns, we apply our image degrading model for LR training data generation instead of an image registration. Practically, we generate 4-times lower-resolution simulations using 20 × realistic images, to constitute the 5 × and 20 × training data pairs with their FOV intrinsically aligned. Vignette high-resolution views of the RFGANM outputs are provided in Figs. 6(c) – 6(f) with a reconstructed pixel size of 0.325 μm. The imaging FOV is around 24 mm^2, the same as that from a × 4 objective (Plan APO, 0.1 NA, Olympus), whereas the maximum achieved resolution is similar to that of a typical 20 × objective (Plan APO, 0.45 NA, Olympus). The conventional microscope images taken with 20 × and × 4 lenses are shown for comparison in Figs. 6(c2) and 6(d2), 6(c3) and 6(d3), respectively. Figures 6(c4) and 6(d4) are the deconvolution results of 6(c3) and 6(d3), by Lucy-Richardson method, which provide no significant resolution promotions at all. It is noteworthy that beside the superior SISR capability, reconstructed image also shows a large depth of field (DOF) which is inherited from the × 4 source measurement. As a result, it behaves even better than the 20 × HR image in regions such as Fig. 6(d2), where 20 × measurement is slightly out of focus due to its relatively small DOF. This underlying robustness of RFGANM indicates its easy implementation on a broad range of samples with natively unsmooth surfaces.
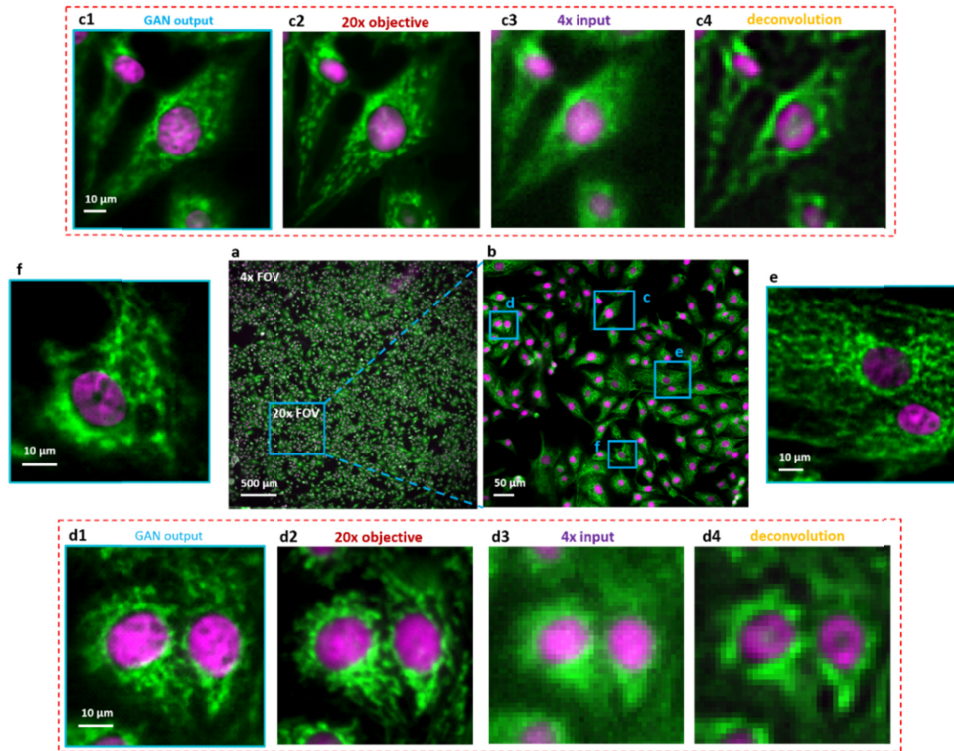
Fig. 6. Dual-color fluorescence imaging via RFGANM. (a) RFGANM-reconstruction of a wide-FOV fluorescence image of BPAE cells specifically labelled with DAPI and Alexa Fluor-488, at nucleus and skeletons, respectively. (a) and (b) show the imaging FOV of 4 × (which RFGANM inherits) and 20 × objectives (high-resolution conventional microscopy), respectively, by using a sCMOS camera (sensor area 1.33 × 1.33 cm). (c1), (d1), (e) and (f) High-resolution views of the selected regions (blue) in (a). (c2) and (d2) High resolution iamges taken under a conventional wide-field fluorescence microscope with a 20 × /0.45 objective. (c3) and (d3) Low resolution inputs taken under 4 × /0.1 objective. (c4) and (d4) The deconvolution results of (c3) and (d3) respectively.

2.6 High-throughput, gigapixel color imaging of pathological slides

Several biomedical studies, such as histology and pathology, intrinsically need the observation of both large-scale statistics of tissue and local details of the cells. Here we further apply RFGANM to imaging of human tissue sections that are from a normal prostate and a prostate tumor. HR reference images were taken by a macro-zoom microscope (Olympus mv10 ×), using 10 × magnification plus a color camera (QHY247C, 3.9 μm pitch size). The corresponding 2.5 × LR training images were down-sampled from the HR measurements via the degrading model. The real 2.5 × measurements of the tissue slides have a large FOV of about 64 mm^2, which almost covers the entire pathological sections (Figs. 7(a) and 7(f)). However, the magnified views show that the high-frequency details are largely decimated (Figs. 7(c1) and 7(i1)), and the deconvolution results by Lucy-Richardson method create artificial details that don't exist in ground truth. The 10 × HR measurements contain abundant structure details, as shown in Figs. 7(c2) and 7(i2), but suffer from much smaller FOV (~4 mm^2, inset in Fig. 7(a)). RFGANM, in contrast, is capable of providing numerous subcellular details across the entire sample scale. Vignette high-resolution views from different areas of the RFGANM reconstruction are shown in Figs. 7(b)-7(e) and Figs. 7(g)-7(j), to verify its strong imaging capability. The RFGANM-reconstructed pathology slides encompass 0.4 gigapixels, with an effective SBP being ~0.1 gigapixels, which is 16-times higher than that of a conventional microscope. It is also noted that the pathological tissue

slides often contain a lot of textures at various scales, thus posing a big challenge to the performance of super-resolution models. Our approach, under this circumstance, still achieves a high-fidelity result with high-frequency textures recovered.
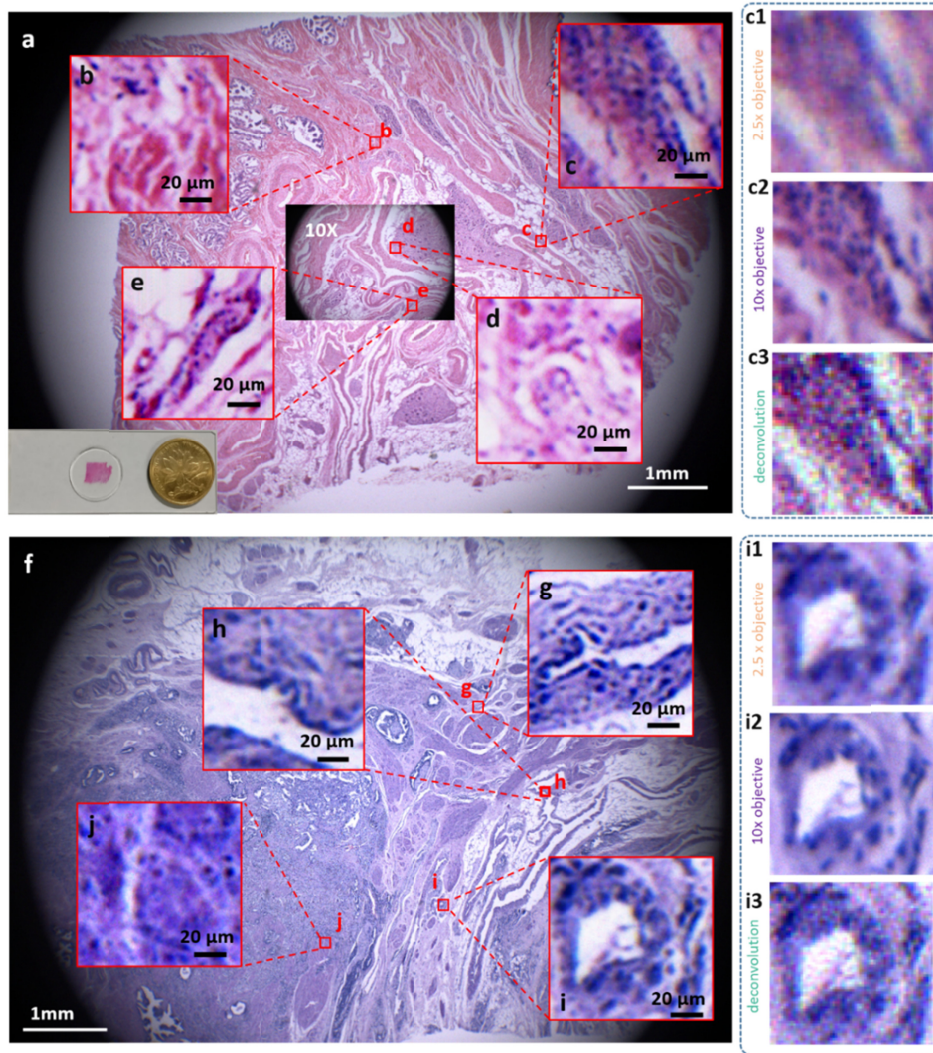

Fig. 7. Gigapixel color imaging of prostate tissue slides. (a) and (f) RFGANM color images of normal prostate histology slide and prostate cancer pathology slide, respectively. The achieved effective SBP here is ~0.1 gigapixels. (b-e) Vignette high-resolution views of the image in (a). (g-j) Vignette high-resolution views of the image in (f) .(c1) and (i1) The 2.5 × input images of RFGANM to get (c) and (i); (c2) and (i2) Images taken by an Olympus macro-zoom microscope under 10 × magnifications; (c3) and (i3) Deconvolution results of 2.5 × inputs (c1) and (i1), respectively, for comparison with RFGANM results.

2.7 RFGANM in light-sheet imaging of selective planes in whole mouse brain

The advent of light-sheet fluorescence microscopy (LSFM) has greatly revolutionized three-dimensional (3-D) imaging of objects from cells to animals [39]. Compared to the epifluorescence methods, such as wide-field and confocal microscopes, LSFM has the advantages of tightly confined excitation, relatively low phototoxicity and high speed [40]. As a reference point, recent integration of LSFM and tissue-clearing has become an important

alternative to conventional histology imaging approaches. However, even for current LSFM, the optical throughput of the system optics remains insufficient to *intoto* map the cellular information throughout a specimen of large volume size; for example, for visualization of the fine neuronal networks across a mouse brain. Tile imaging is the commonly-used approach to artificially increase the SBP, and realize high-resolution imaging of large specimens [41,42]. Despite the compromised speed induced by repetitive mechanical stitching, the high illumination/ detection NA configuration in tile imaging greatly limits the fluorescence extraction from deep tissue of the thick specimens. We demonstrate, instead of commonly-used tile stitching, which significantly sacrifices the throughput and limit the signal extraction from deep tissue, the integration of RFGANM with light-sheet imaging can achieve high-resolution imaging of selective sectional planes in a whole adult mouse brain. We first constructed a macro-view light-sheet geometry with wide laser-sheet illumination and large-FOV detection (Fig. 8(a)), which can fully cover an optically-cleared P30 mouse brain (Tg: Thy1-GFP-M). 200 consecutively-illuminated transverse planes in the middle of the brain (depth 2 to 3 mm) were recorded (Fig. 8(b)), with their maximum-intensity-projection (MIPs) showing the global distribution of the neurons. The raw plane images simply accept the limited resolution from the macro-view LSFM system optics, hence the densely-packed neuronal fibers remain dim. The super-resolved image is then instantly obtained by RFGANM, with a reconstructed pixel size of 0.53 μm (Figs. 8(c)-8(h)). The -result is furthermore compared to higher-magnification light-sheet measurements (6.4 × detection) to confirm the authenticity of the computation. In Fig. 8(d3), the neuronal dendrites identified by each method reveal substantially improved resolution from RFGANM. Therefore, besides conventional epifluorescence methods, RFGANM is proven to be the same efficient to the LSFM imaging, which together are capable of rapidly obtaining the high-resolution signals from arbitrary planes of intact large tissues. Furthermore, in the light of the strong 3-D imaging capability of LSFM, RFGAN-LSFM is possibly to be extended to the third dimension in the future, to achieve high-throughput, high-resolution volumetric mapping of whole specimens, such us intact organs, and whole embryos.
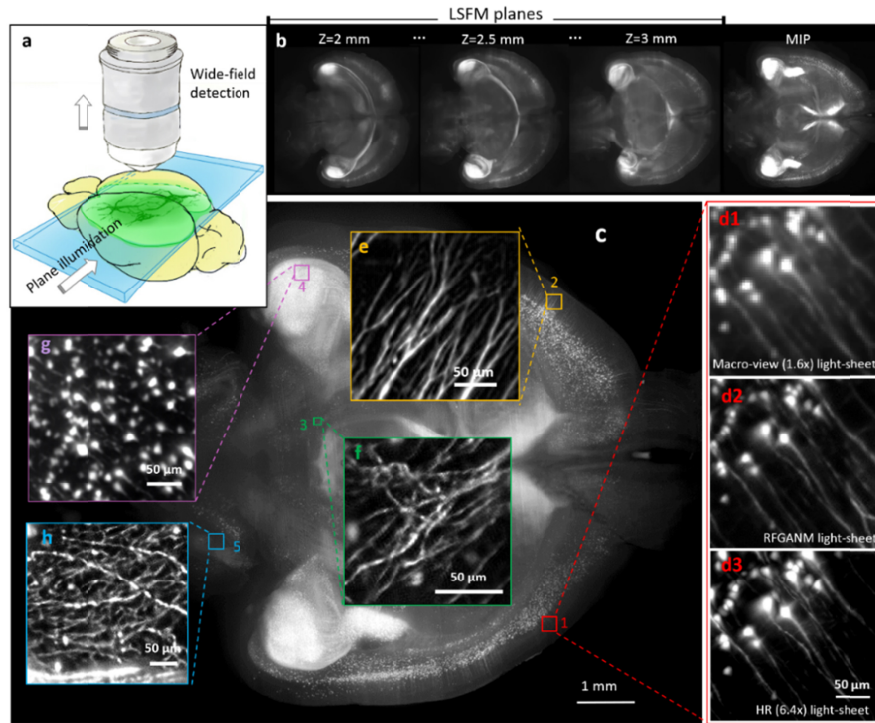
Fig. 8. GAN-based light-sheet fluorescence microscopy. (a) Schematic of light-sheet imaging geometry for optical sectioning in the deep of intact organ. (b) High-contrast plane images of the mouse brain. (c) A 1mm-thick transverse slice of the whole brain, which is the MIP of 200 consecutively illuminated planes. The image was then super-resolved with 4 times enhancement, and compared to raw 1.6 × (d1) and 6.4 × measurements (d3). (e-h) Magnified reconstruction views of selective regions of cortex, telencephalon, hippocampus, and cerebellum.

2.8 Quantitative evaluation of image similarity

We calculate the peak signal-to-noise ratio (PSNR[dB]) and the structural similarity (SSIM) [43] index between the RFGANM outputs and the high-resolution measurements of the fluorescent BAEP cells and pathological section (Table 1), to quantify the accuracy of the RFGANM. For comparison, PSNR and SSIM index between the bicubic interpolation followed by deconvolution of the low-resolution measurements and the corresponding high-resolution measurements are also computed. As a reference, the PSNR and SSIM of the high-resolution measurement to itself is ∞ and 1 respectively. Each test group contains 16 samples with a size of 384*384 pixels. We compute their PSNR and SSIM respectively and take the averaged value of all the 16 samples as final results. Apparently, SSIM and PSNR of the network reconstructions are both far better than that of the deconvolution results of bicubic interpolation of the LR measurements. The increase of PSNR from 19.63dB to 27.26dB validates the remarkable resolution enhancement, meanwhile the high-level SSIM index of 93.17% proves the authenticity of reconstruction.

Table 1. PSNR and SSIM between the GAN-reconstructed results and the realistic HR measurements

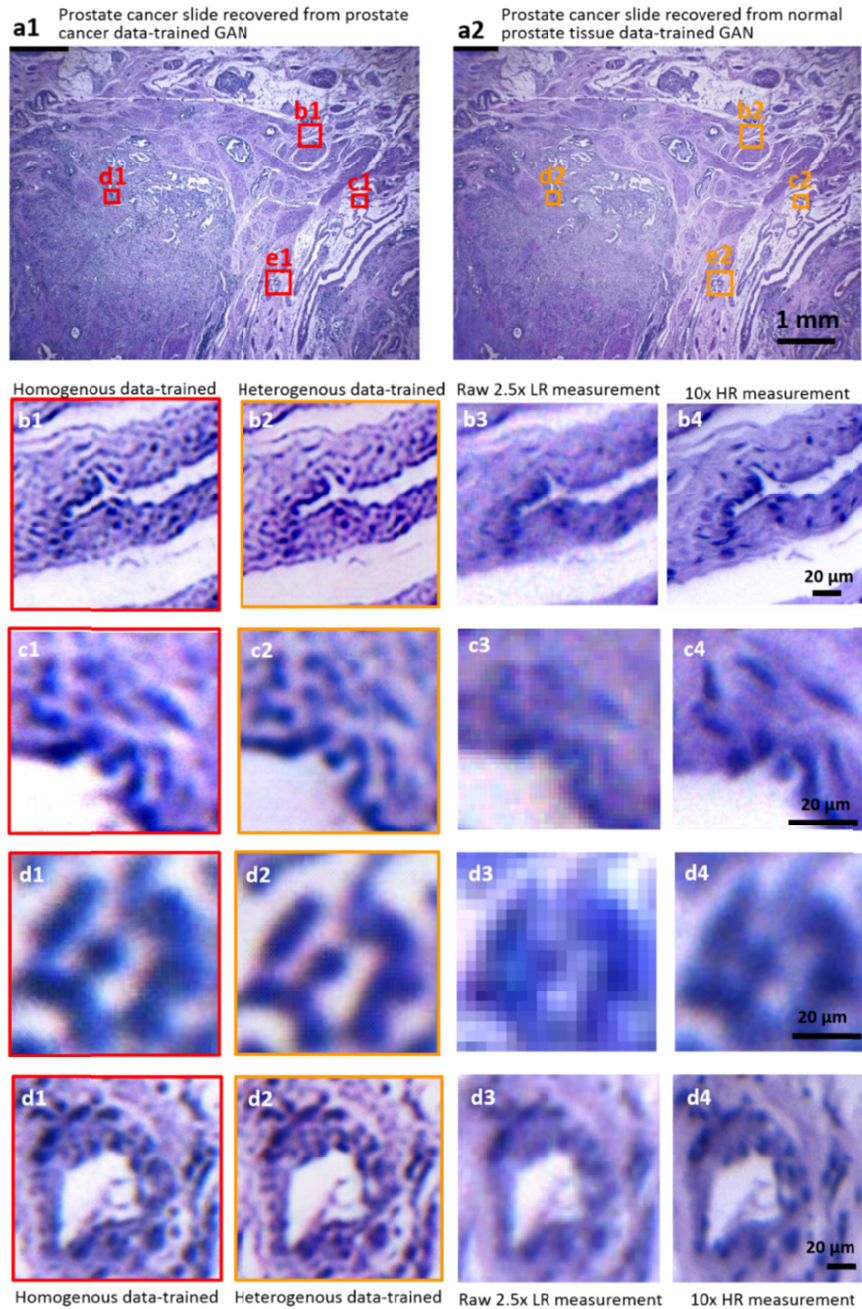| PSNR/SSIM	BPAE	Pathological section
Deconvolution of bicubic interpolation	19.63/0.81	24.36/0.91
Network reconstruction	27.79/0.93	25.60/0.92
High resolution Target	∞/1	∞/1

2.9 The robustness of the network

Fig. 9. Validating the robustness of RFGANM. (a1) and (a2) Wide-FOV reconstruction images of a prostate cancer slide using homogenous-data-trained (prostate cancer images) network and heterogenous-data-trained (healthy prostate images) network. (b1-d1) Vignette high-resolution views of the image in (a1). (b2-d2) Vignette high-resolution views of the (a2). (b3-d3) Images taken by a macro-zoom microscope (mv10 ×) under 2.5 × magnification. (b4-d4) Images taken under 10 × magnification.

The training process of the GAN network is fundamentally a process in which the neural network extracts features from the training data and learns their roles in the corresponding LR

and HR counterparts. Theoretically a network trained by one type of sample images should be also applicable to similar types of samples. For example, we can reasonably speculate that a GAN generator well-trained by healthy prostate data can work with the prostate cancer tissue as well. To test this underlying robustness, we blindly apply the network trained by healthy prostate tissue images to reconstruct a low-resolution image of prostate cancer tissue. Its outputs are compared with those from a real prostate-cancer-data-trained network, as shown in Fig. 9 below. Both networks recover highly similar structures with similar qualities presented, capable of resolving high-resolution cellular details, such as the nucleus and textures. It strongly suggests that GAN network could be highly robust, implying that RFGANM can go further with being applied to the reconstruction of a variety of samples merely with single type of data training.

2.10 Histopathological diagnosis and cell counting by RFGANM

Large-scale quantitative analyses are further enabled based on RFGANM imaging, as shown in Fig. 10. In the four segmented encephalic regions of a whole mouse brain section (Fig. 10(a)), the neurons are identified and calculated with populations (Imaris visualization software, Fig. 10(b)) using $1.6 \times$ LR, $6.4 \times$ HR and RFGANM images, respectively. The counting results are plotted as Fig. 10(c). Due to a severe structure details decimation, $1.6 \times$ LR results fail to precisely count the numbers of neurons recognized. Even denoised using a widely accepted BD3M algorithm [44,45], the results of $1.6 \times$ are still far away from HR ground truth. In contrast, the counted numbers of neurons in RFGANM images are very close to those of the $6.4 \times$ HR measurements. Figures 10(d)-10(i) further show the cell nuclei counting of the healthy prostate tissue and the prostate cancer tissue images. Figure 10(d) shows the normal prostate gland, which has a lobular architecture. The glands are grouped and often have folded contour. Figure 10(f) shows the glands of prostate cancer invasion, which has predominant cribriform glands, lack of the component of well-formed glands. Unlike the raw low-resolution measurement which is not able to discern the single cell nuclei in detail, GANM image here enables fairly accurate cell number counting within the micro-tumor, which is beneficial to the doctor for more specific rating of the cancer invasion. With the integration of RFGANM into biomedical application, which is featured by the combination of sing-cell resolution and centimeter large field view at seconds high throughput, biological research or clinical work efficiency can be considerably improved for quantitatively checking of large mass surgical specimens.
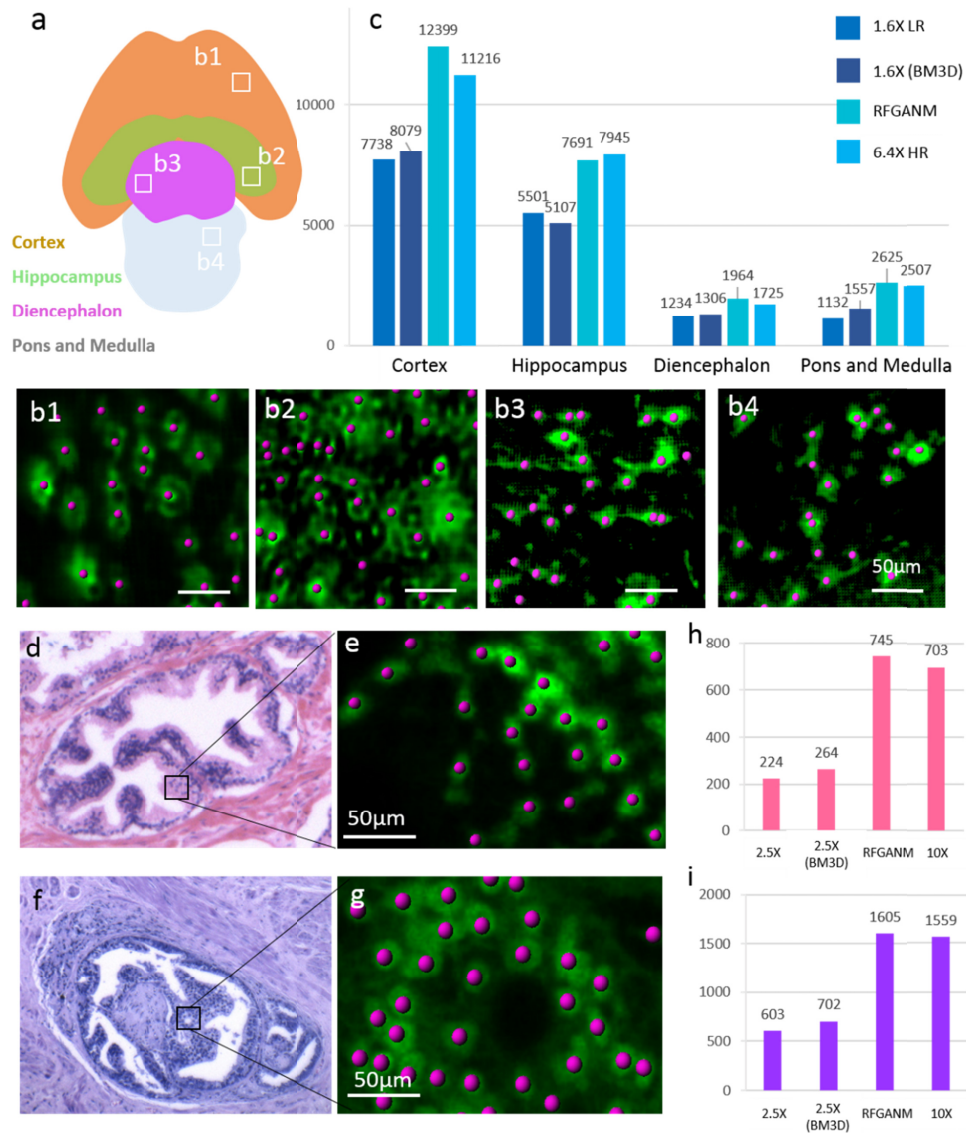
Fig. 10. Biomedical quantitative analysis based on RFGANM images. (a) The encephalic regions division of the whole mouse brain section. 4 main encephalic regions are segmented as Cortex, Hippocampus, Diencephalon, Pons and Medulla. (b1-b4) Examples of identified and counted neurons in 4 regions, respectively. (c) The calculated cellular populations at different brain regions of (a). The counting results of RFGAM are consistent with those from HR measurements. (d and f) The selected gland areas from RFGANM images of a prostate tissue and a prostate cancer slide, respectively, for cell nuclei counting. (e and g) Examples of identified and counted cell nuclei. (h and i) The counting results of (d) and (f), respectively.

3. Methods

3.1 Sample preparation

The fluorescent images in our experiments were taken on bovine pulmonary artery endothelial cells (BPAEC) that were fixed and multi-labeled before imaging. In cell preparation, after cell fixation and permeabilization, F-actin of cells was stained by Alexa Fluor 488 phalloidin (Thermo Fisher Scientific Inc.), and the nuclei were counterstained using

the blue-fluorescent DNA stain DAPI. The pathological tissue slides for bright-field microscopy were healthy human prostate tissue and human prostate cancer tissue, respectively, stained with hematoxylin-eosin after paraffin-embedding and automatic slicing. For light sheet imaging, an uDISCO cleared P30 mouse brain is used.

3.2 Training implementations

Our model is implemented based on Google's deep learning framework, TensorFlow (version r1.8), and trained on an Inspur server with two NVidia Tesla P100 GPUs. The source code together with a small example data set is available on GitHub (https://github.com/xinDW/RFGANM). Initiating with a batch size of 16 and a fixed learning rate of 10^{-4}, we trained the network for 200 epochs, which took about 48 hours.

3.3 Inference process

In the inference phase after network training, the experimentally captured LR images for validation are cut into a bunch of small pieces with overlaps with each other, and then input into the network for super-resolved reconstruction piece by piece. Afterwards, all these output pieces are stitched into one whole image that possesses both large FOV and high resolution. The stitching process is achieved by matching the overlapped regions, which is very robust and accurate. The inference process is quite fast. An image piece of 100×100 pixels size takes less than 0.01 second to be super-resolved into a 400×400 pixels image, even on an ordinary Windows laptop with Intel Core i5 CPU.

3.4 Imaging setups

There are several kinds of microscopy images in our experiment: the bright-field grayscale resolution test target images, the dual-color fluorescent BPAE images, the bright-field color images of two types of tissue slides, and the light sheet images of mouse brain. Images of resolution target were recorded by a Photometrics IRIS15 camera (pitch size 4.25 μm), with the HR and LR images taken under $10 \times$ and $2 \times$ magnifications of an Olympus MVX10 microscope, respectively. The BPAE cells were imaged under an Olympus IX73 microscope equipped with a HAMAMATSU ORCA-Flash 4.0-V2 camera (pitch size 4.25 μm). In both fluorescent channels, the HR training images and LR validation images were taken under a 2 $\times 20 \times /0.45$ and $4 \times /0.1$ objective, respectively. For pathology slide imaging, a QHY247C color camera (pitch size is 3.9 μm) was used on the Olympus MVX 10 microscope to capture the healthy prostate/prostate cancer tissues stained with hematoxylin-eosin. The HR training and LR validation images were then taken under $10 \times$ and $2.5 \times$ magnifications, respectively. The sectional images of intact mouse brain were obtained by a macro-view light-sheet system, which comprised a self-made dual-side laser-sheet illumination and large-FOV wide-field detection by Olympus MVX10 microscope body. The images were sequentially recorded using Photometrics IRIS15 camera under $1.6 \times$ and $6.4 \times$ magnifications.

4. Conclusion

We have demonstrated a deep learning-based microscopy method without the requirement of extra registration procedure in training course, which can significantly improve the resolution of conventional wide-field and cutting-edge light-sheet fluorescence microscopes, and greatly increase the imaging throughput for whole biomedical specimens. We apply a state-of-the-art GAN network to deeply learn how to map from the low-resolution microscopy images to their high-resolution counterparts. For cell and tissue images that contain complicated patterns, their low-resolution training data are artificially generated and intrinsically registered to the high-resolution training images via a degradation model. This step has simplified the data preprocessing and improved the robustness of the GAN network. Once the model training being accomplished, the well-established AI agent is capable of quickly reconstructing a large FOV, super-resolution image of new sample based on a single low-resolution snapshot taken

by an ordinary optical microscope. Besides the improved resolution that has been verified by imaging of resolution target and PSNR analysis, the structure similarity to the sample ground truth has also been quantified, at a high level of over 90%. Currently we have proven RFGANM method could be beneficial to biomedical applications such as cell counting and histopathological diagnoses. At the same time, the artifacts existing in some minor regions still need to be refined in future work, by upgrading the network structure as well as the training algorithm. We also prove that this RFGANM method is very robust, readily applicable to most forms of microscopy data such as bright-field images, epifluorescence images, and light-sheet fluorescence images. It significantly extends the SBP of these microscope systems neither at the cost of acquiring multiple frames nor relying on the retrofit of conventional microscope system. Therefore, RFGANM has a high temporal performance, but shows a much better image quality that is comparable to those multi-frame SR methods. As a reference point, it produces a 0.38 gigapixel digital pathology slide at 1 μm resolution, with an acquisition time of 0.01 second and computation time of less than 1 second. This high-resolution combined with high-throughput capability renders RFGANM a valuable tool for many applications, such as tissue pathology and neuroanatomy. Furthermore, though currently we demonstrate the combination of deep learning and convolutional neural network with optical microscopy in form of 2-D imaging of *exvivo* samples, we can reasonably expect that provided its superior spatial-temporal performance, this methodology will be also applicable to both 3-D microscopy and highly dynamic process.

Funding

Acknowledgements

Disclosures

The authors declare that there are no conflicts of interest related to this article.

References

1. G. Zheng, R. Horstmeyer, and C. Yang, "Wide-field, high-resolution Fourier ptychographic microscopy," Nat. Photonics **7**(9), 739–745 (2013).
2. A. W. Lohmann, R. G. Dorsch, D. Mendlovic, C. Ferreira, and Z. Zalevsky, "Space–bandwidth product of optical signals and systems," J. Opt. Soc. Am. A **13**(3), 470–473 (1996).
3. G. Hilgen, S. Pirmoradian, A. Maccione, L. Berdondini, M. H. Hennig, and E. Sernagor, "High Resolution Large-Scale Recordings of Light Responses from Ganglion Cells in the Developing Mouse Retina," in *FASEB Conference on Retinal Neurobiology and Visual Processing*(2014).
4. K. Imfeld, S. Neukom, A. Maccione, Y. Bornat, S. Martinoia, P. A. Farine, M. Koudelka-Hep, and L. Berdondini, "Large-scale, high-resolution data acquisition system for extracellular recording of electrophysiological activity," IEEE Trans. Biomed. Eng. **55**(8), 2064–2073 (2008).
5. K. Imfeld, S. Neukom, A. Maccione, Y. Bornat, S. Martinoia, P. A. Farine, M. Koudelka-Hep, and L. Berdondini, "Large-scale, high-resolution data acquisition system for extracellular recording of electrophysiological activity," IEEE Trans. Biomed. Eng. **55**(8), 2064–2073 (2008).
6. M. Brown and D. G. Lowe, "Automatic panoramic image stitching using invariant features," Int. J. Comput. Vis. **74**(1), 59–73 (2007).
7. T. R. Hillman, T. Gutzler, S. A. Alexandrov, and D. D. Sampson, "High-resolution, wide-field object reconstruction with synthetic aperture Fourier holographic optical microscopy," Opt. Express **17**(10), 7873–7892 (2009).
8. M. Kim, Y. Choi, C. Fang-Yen, Y. Sung, R. R. Dasari, M. S. Feld, and W. Choi, "High-speed synthetic aperture microscopy for live cell imaging," Opt. Lett. **36**(2), 148–150 (2011).

9.   T. Gutzler, T. R. Hillman, S. A. Alexandrov, and D. D. Sampson, "Coherent aperture-synthesis, wide-field, high-resolution holographic microscopy of biological tissue," Opt. Lett. **35**(8), 1136–1138 (2010).

10. W. Luo, A. Greenbaum, Y. Zhang, and A. Ozcan, "Synthetic aperture-based on-chip microscopy," Light Sci. Appl. **4**(3), e261 (2015).

11. G. Zheng, S. A. Lee, S. Yang, and C. Yang, "Sub-pixel resolving optofluidic microscope for on-chip cell imaging," Lab Chip **10**(22), 3125–3129 (2010).

12. G. Zheng, S. A. Lee, Y. Antebi, M. B. Elowitz, and C. Yang, "The ePetri dish, an on-chip cell imaging platform based on subpixel perspective sweeping microscopy (SPSM)," Proc. Natl. Acad. Sci. U.S.A. **108**(41), 16889–16894 (2011).

13. W. Luo, Y. Zhang, A. Feizi, Z. Göröcs, and A. Ozcan, "Pixel super-resolution using wavelength scanning," Light Sci. Appl. **5**(4), e16060 (2016).

14. W. Xu, M. H. Jericho, I. A. Meinertzhagen, and H. J. Kreuzer, "Digital in-line holography for biological applications," Proc. Natl. Acad. Sci. U.S.A. **98**(20), 11301–11305 (2001).

15. L. Denis, D. Lorenz, E. Thiébaut, C. Fournier, and D. Trede, "Inline hologram reconstruction with sparsity constraints," Opt. Lett. **34**(22), 3475–3477 (2009).

16. A. Greenbaum, W. Luo, B. Khademhosseinieh, T. Su, A. F. Coskun, and A.Ozcan, "Increased space-bandwidth product in pixel super-resolved lensfree on-chip microscopy," Sci. Rep. **3**(1), 1717 (2013).

17. M. Elad and Y. Hel-Or, "A fast super-resolution reconstruction algorithm for pure translational motion and common space-invariant blur," IEEE Trans. Image Process. **10**(8), 1187–1193 (2001).

18. S. Farsiu, M. D. Robinson, M. Elad, and P. Milanfar, "Fast and robust multiframe super resolution," IEEE Trans. Image Process. **13**(10), 1327–1344 (2004).

19. P. Vandewalle, S. Sü, and M. Vetterli, "A frequency domain approach to registration of aliased images with application to super-resolution," EURASIP J. Adv. Signal Process. **2006**, 1–14 (2006).

20. M. G. L. Gustafsson, "Surpassing the lateral resolution limit by a factor of two using structured illumination microscopy," J. Microsc. **198**(Pt 2), 82–87 (2000).

21. M. G. Gustafsson, "Nonlinear structured-illumination microscopy: wide-field fluorescence imaging with theoretically unlimited resolution," Proc. Natl. Acad. Sci. U.S.A. **102**(37), 13081–13086 (2005).

22. J. Yang, J. Wright, T. Huang, and Y. Ma, "Image super-resolution as sparse representation of raw image patches," in *IEEE Conference on Computer Vision and Pattern Recognition* (2008), pp. 1–8.

23. R. Timofte, V. D. Smet, and L. V. Gool, *A+: Adjusted Anchored Neighborhood Regression for Fast Super-Resolution* (Springer International Publishing, 2014).

24. M. G. Ertosun and D. L. Rubin, "Automated Grading of Gliomas using Deep Learning in Digital Pathology Images: A modular approach with ensemble of convolutional neural networks," AMIA Symposium **2015**, 1899 (2015).

25. A. Cruzroa, A. Basavanhally, H. Gilmore, M. Feldman, S. Ganesan, N. Shih, J. Tomaszewski, and A. Madabhushi, "Automatic detection of invasive ductal carcinoma in whole slide images with convolutional neural networks," Proceedings of SPIE - The International Society for Optical Engineering **9041**, 139–144 (2014).

26. G. Litjens, C. I. Sánchez, N. Timofeeva, M. Hermsen, I. Nagtegaal, I. Kovacs, C. Hulsbergen-van de Kaa, P. Bult, B. van Ginneken, and J. van der Laak, "Deep learning as a tool for increased accuracy and efficiency of histopathological diagnosis," Sci. Rep. **6**(1), 26286 (2016).

27. A. Ozcan, H. Günaydin, H. Wang, Y. Rivenson, Y. Zhang, and Z. Göröcs, "Deep learning microscopy," Optica **4**, 1437–1443 (2017).

28. Y. Rivenson, H. C. Koydemir, H. Wang, Z. Wei, Z. Ren, H. Gunaydin, Y. Zhang, Z. Gorocs, K. Liang, and D. Tseng, "Deep learning enhanced mobile-phone microscopy," ACS Photonics **5**, 2354–2364 (2017).

29. W. Ouyang, A. Aristov, M. Lelek, X. Hao, and C. Zimmer, "Deep learning massively accelerates super-resolution localization microscopy," Nat. Biotechnol. **36**(5), 460–468 (2018).

30. E. Nehme, L. E. Weiss, T. Michaeli, and Y. Shechtman, "Deep-STORM: super-resolution single-molecule microscopy by deep learning," Optica **5**, 458–464 (2018).

31. N. Boyd, E. Jonas, and H. B. B. Recht, "DeepLoco: fast 3D localization microscopy using neural networks," http://bioRxiv 236463 (2018).

32. Y. R. Hongda Wang, Y. Jin, Z. Wei, R. Gao, Harun, L. A. B. Günaydın, and A. Ozcan, "Deep learning achieves super-resolution in fluorescence microscopy," https://bioRxiv 309641 (2018).

33. H. Zhang, X. Xie, C. Fang, Y. Yang, D. Jin, and P. Fei, "High-throughput, high-resolution Generated Adversarial Network Microscopy," https://arxiv.org/abs/1801.07330 (2018).

34. I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *International Conference on Neural Information Processing Systems*(2014), pp. 2672–2680.

35. W. Shi, J. Caballero, F. Huszar, J. Totz, A. P. Aitken, R. Bishop, D. Rueckert, and Z. Wang, "Real-time single image and video super-resolution using an efficient sub-pixel convolutional," Neural Netw. **2016**, 1874–1883 (2016).

36. C. Ledig, Z. Wang, W. Shi, L. Theis, F. Huszar, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, and A. Tejani, "Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network," https://arxiv.org/abs/1609.04802 (2016).

37. J. Johnson, A. Alahi, and L. Fei-Fei, "Perceptual Losses for Real-Time Style Transfer and Super-Resolution," in *European Conference on Computer Vision* (2016), pp. 694–711.

38. K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," Comput. Sci. (2014).
39. J. Huisken, J. Swoger, J. Wittbrodt, and E. H. K. Stelzer, "Optical sectioning deep inside live embryos by selective plane illumination microscopy," Science **305**, 1007–1009 (2004).
40. P. J. Keller, A. D. Schmidt, J. Wittbrodt, and E. H. K. Stelzer, "Reconstruction of Zebrafish Early Embryonic Development by Scanned Light Sheet Microscopy," Science **322**(5904), 1065–1069 (2008).
41. K. Becker, N. Jährling, E. R. Kramer, F. Schnorrer, and H. U. Dodt, "Ultramicroscopy: 3D reconstruction of large microscopical specimens," J. Biophotonics **1**(1), 36–42 (2008).
42. Z. Guan, J. Lee, H. Jiang, S. Dong, N. Jen, T. Hsiai, C. M. Ho, and P. Fei, "Compact plane illumination plugin device to enable light sheet fluorescence imaging of multi-cellular organisms on an inverted wide-field microscope," Biomed. Opt. Express **7**(1), 194–208 (20165).
43. B. A. C. Wang Z., H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," IEEE Trans. Image Process. **13**(4), 13 (2004).
44. K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising by sparse 3-D transform-domain collaborative filtering," IEEE Trans. Image Process. **16**(8), 2080–2095 (2007).
45. K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Color Image Denoising via Sparse 3D Collaborative Filtering with Grouping Constraint in Luminance-Chrominance Space," in *IEEE International Conference on Image Processing*(2007), pp. 313–316.