

# High-throughput, high-resolution Generated Adversarial Network Microscopy

Hao Zhang<sup>1+</sup>, Xinlin Xie<sup>1+</sup>, Chunyu Fang<sup>1</sup>, Yicong Yang<sup>1</sup>, Di Jin<sup>3\*</sup> and Peng Fei<sup>1,2\*</sup>

<sup>1</sup> School of Optical and Electronic Information, Huazhong University of Science and Technology, Wuhan, 430074, China.

<sup>2</sup> Britton Chance Center for Biomedical Photonics, Wuhan National Laboratory for Optoelectronics, Huazhong University of Science and Technology, Wuhan, 430074, China.

<sup>3</sup> Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge, MA, 02139, U.S.A.

\*Correspondence: [feipeng@hust.edu.cn](mailto:feipeng@hust.edu.cn)

<sup>+</sup> Equal contributing authors

## Abstract

We for the first time combine generated adversarial network (GAN) with wide-field 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 network can recover a high-resolution, accurate image of new specimen from its single low-resolution measurement. This capacity has been adequately demonstrated by imaging various types of samples, such as USAF resolution target, human pathological slides and fluorescence-labelled fibroblast cells. Their gigapixel, multi-color reconstructions verify a successful GAN-based single image super-resolution procedure. Furthermore, this deep learning-based imaging approach doesn't necessarily introduce any change to the setup of a conventional wide-field microscope, reconstructing large FOV ( $\sim 95 \text{ mm}^2$ ), high-resolution ( $\sim 1.7 \text{ }\mu\text{m}$ ) image at a high speed (in 1 second). As a result, GAN-microscopy opens a new way to computationally overcome the general challenge of high-throughput, high-resolution microscopy that is originally coupled to the physical limitation of system's optics.

## Introduction

The imaging throughput of a conventional optical microscope is typically limited to megapixels, regardless of the magnification and numerical aperture used<sup>1,2</sup>. As a result, compromise often exists between achieving a high resolution and maintaining a large field-of-view. 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. 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 followed and quantitatively analyzed at the same time<sup>3</sup>. Development of sophisticated mechanical scanning

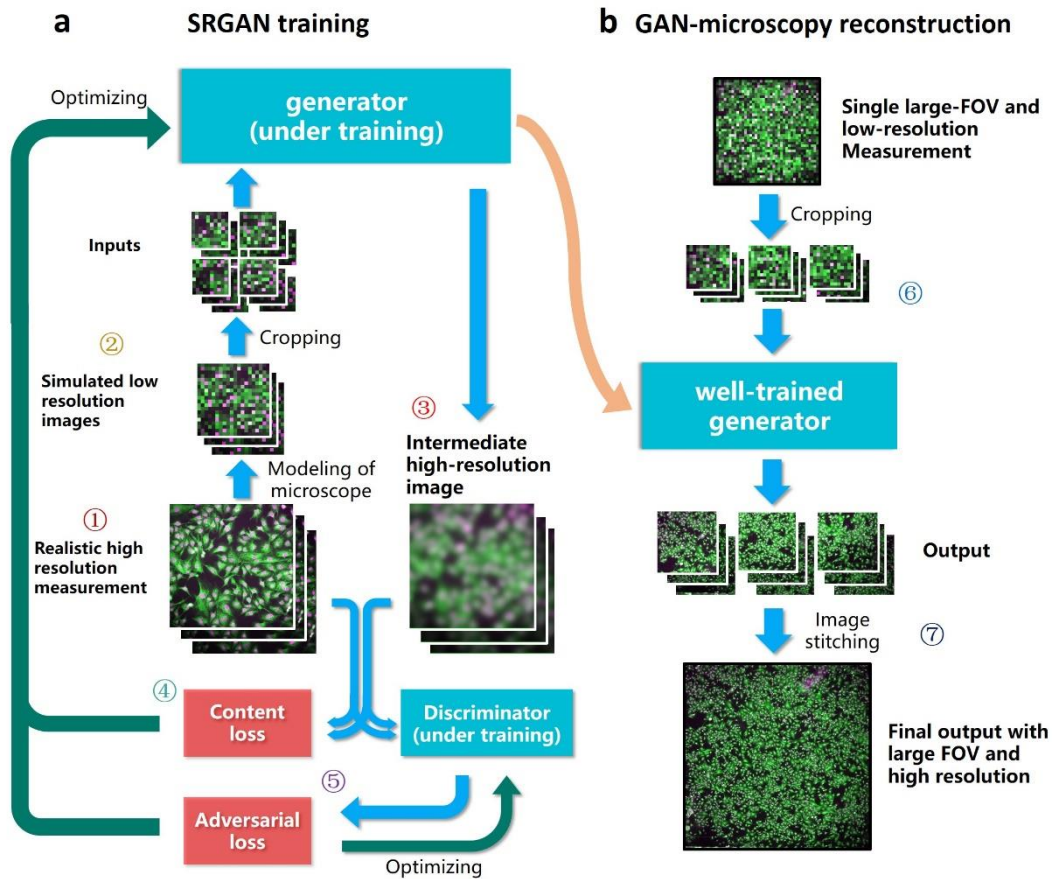
microscope is a commonly-used way to address this challenge, artificially increasing the throughput of the microscope by stitching multiple high-resolution tiles into a panoramic image<sup>4</sup>. Besides this mechanical means that requires precise control over actuation and optical alignment, recent super Resolution (SR) techniques present a computational way to increase the space-bandwidth product of a microscope platform<sup>1,5-17</sup>. 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 shift into a high resolution image<sup>15,16</sup>. Several frequency domain methods, e.g., Fourier ptychographic microscopy (FPM)<sup>1</sup>, Synthetic aperture microscopy<sup>5-8</sup> and Structured-illumination microscopy<sup>18,19</sup>, produce a resolution-enhanced image by stitching together a number of variably illuminated, low-resolution images in Fourier space. Despite offering unique imaging capabilities with scalable SBP, these methods, however, all require special hardware setup and complex computation on multiple frames. Another type of techniques that have been widely applied in microscopy is single image super resolution (SISR). SISR aims at the reconstruction of a high-resolution (HR) images with rich details from single low-resolution (LR) image, usually by means of example-based approach<sup>20,21</sup>. It works by searching the high-resolution patches in the example dictionary and replacing the low-resolution information with the high-resolution ones. Although SISR greatly simplifies the imaging hardware architecture and the computation complexity, 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 efficient SISR. Besides the well-known successes in machine-based segmentation, classification, and even diagnosis<sup>22-27</sup>, it has been reported that the convolutional neural networks (CNNs) can also contribute to resolution improvement of pictures<sup>28-30</sup>. As a demonstration, a 37-layer SRRESNET was successfully developed to super-resolve simulated low-resolution photographs of animals, landscapes and human faces<sup>28</sup> at an improvement factor of 4. Most recent SRGAN with same 37-layer network has claimed even better performance without increasing the training intensity<sup>28</sup>. These successes in regular picture's improvement shows CNNs could also be a promising tool for microscopy, in which the targets are always more unusual and unpredictable.

Here we present a deep learning-based imaging approach that is capable of providing significant resolution improvement for conventional microscope, without the need of acquiring a plurality of frames or retrofitting existing optical systems. This imaging method uses dataset that consists of high-resolution measurements and their low-resolution pairs to train a Generated Adversarial Network (GAN). 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 with prior data is accomplished, the validation 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 GAN-microscopy (GANM) method with bright-field image of USAF resolution target, color image of pathological slides, and dual-channel fluorescence image of fibroblast cells, verifying that it's widely applicable to various wide-field microscopy data. The achieved improvement factor is currently 5, indicating a 25 times higher throughput than the original microscopes. By taking a few example images as the references and applying a GAN deep-learning procedure, we

transform a conventional optical microscope into a high-resolution ( $\sim 1.7\ \mu\text{m}$ ), wide-FOV ( $\sim 95\ \text{mm}^2$ ) microscope with a final SBP of 0.13 gigapixels. Furthermore, the reconstruction procedure is independent from the GPU-based training, thus can be performed on a normal computer. This underlying advantage renders GAN-Microscopy a robust platform that allows multiple applications to be followed once after a well-trained SR artificial intelligence is established. In the following, we will briefly describe the GAN-Microscopy operation and experimental set-up, discuss how to apply the network training and validation process, and discuss its imaging applications in a variety of biomedical samples.

## 1. Results

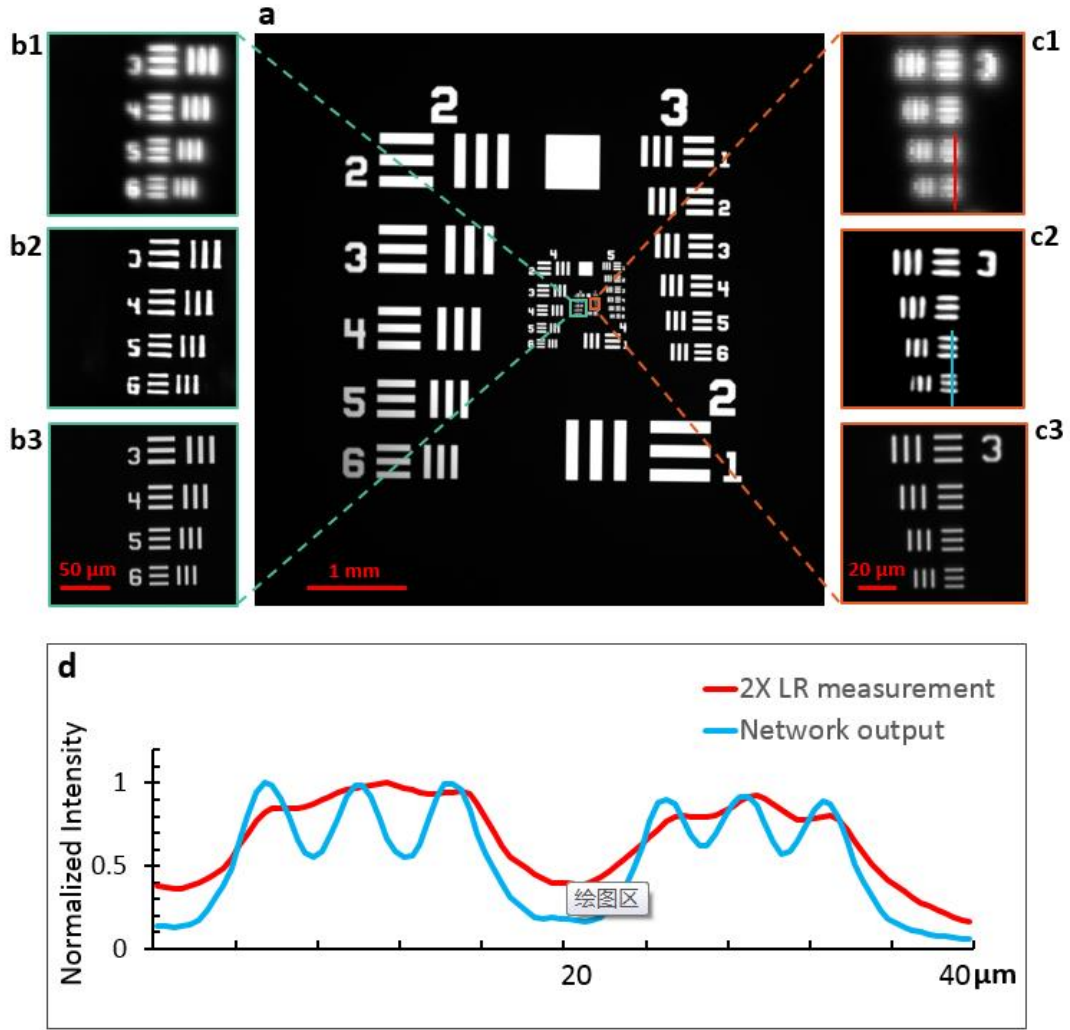
A state-of-the-art convolutional neural network (CNN) architecture, Generative Adversarial Network (GAN)<sup>31</sup>, is used to “learn” the various types of microscopy data from scratch. Figure 1 illustrates the network training and inference process. The GAN majorly contains a generator that takes the prior high-resolution measurements and their degraded low-resolution versions as the inputs to iteratively train the network on how to produce an perceptually high-resolution estimate (shown as Fig. 1a). We first take multiple high-resolution images of the example specimen using high-magnification objective (Fig. 1a, step1). Then through accurately modeling the transfer function of the microscope system in the data preprocessing step, GAN can obtain the down-sampled, blurred images of the example specimen directly via simulation (Fig. 1a, step 2). Based on its currently-learned knowledge, the generator creates resolution-enhanced intermediate outputs in each training iteration (Fig. 1a, step 3). The differences between the intermediate outputs and the realistic high-resolution specimens are calculated, defined as the content loss function of the generator (Fig. 1a, step 4). Unlike other CNN networks, GAN includes an additional discriminator that aims to evaluate the reliability of the generator. The discriminator makes a judgement on whether an image is a guess of the generator or a realistic high-resolution measurement, after they are randomly inputted (Fig. 1a, 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 can be terminated when the generator produces results that the discriminator can hardly tell from the realistic HR images. Then a low-resolution measurement of sample, which is distinct from the data used in the training network, is divided into several patches and fed into the well-trained generator (Fig. 1b, step 6). The generator is capable of recovering high frequency information for each patch, based on its prior GAN training. These quality-improved patches are finally stitched into one gigapixel image of sample that encompasses high resolution details as well as large field of view (Fig. 1b, step 7). The aforementioned image reconstruction process, also termed as inference, is illustrated in figure 1b. The details about GAN architecture and the modeling of microscope degradation are further elucidated in Methods and Supplementary Materials.



**Figure 1. Principle of GAN-microscopy procedure.** A modified Generative Adversarial Network (GAN) is used to iteratively learn the microscopy data till the goal of high-quality output is reached. **a**, The architecture of the GAN training network. It majorly consists of a generator and a discriminator. The generator takes the low resolution, simulated images (step 1, 2) as the inputs and produces their resolution-enhanced versions based on its currently-learned “knowledge” (step 3). A content loss function, defined as the weighted sum of MSE loss and feature reconstruction loss between HR measurements and generator’s intermediate guess (Supplementary Note 1.1), is used to iteratively train the generator towards the generation of accurate outputs that are similar enough with the HR measurements (step 4). A discriminator is designed to determine whether a randomly inputted image is a realistic high-resolution measurement or just a guess of the generator. The probabilities of the discriminator making incorrect decisions (Supplementary Note 1.2) are defined as the adversarial loss, which is further fed back to the discriminator to push it towards higher accuracy on discerning the generator’s outputs from HR ground truths (step 5). This dynamic competition between the generator and discriminator forms an adversarial neural Network. **b**, The GAN-microscopy reconstruction. First a large-FOV, low-resolution measurement of new sample, meaning data not included in previous training, is subdivided into a number of small tiles before being imported into the well-trained GAN network. The network then quickly generates super-resolved images of these small tiles (step 6) and stitches them together into a high-resolution, panoramic image (step 7). GAN-microscopy is capable of providing a gigapixel image at a high speed of single shot, showing good potentials for high-resolution, high-throughput mapping of large specimens.

## 1.1 Characterization of GAN-microscopy

The capability of GAN-microscopy is first characterized through imaging a negative USAF resolution target (Thorlabs R3L3S1N) with highest resolution of 228 line pairs/mm. We capture HR and corresponding LR images under a macro microscope (Olympus MVX10) with  $\times 10$  and  $\times 2$  total magnifications, respectively. An image registration is applied to match their corresponding FOV, forming strictly aligned HR and LR pairs for the GAN training. Considering only simple patterns presented in the resolution target, we apply a geometric transformation, such as translation and rotation, to these paired images to further expand the scale of the dataset. Finally, over 1000 groups of images are imported into the GAN network for training. Another large FOV, LR measurement was used to validate the converged network (Fig. 2a). As shown in Fig. 2b to 2d, 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  $\times 2$  image can hardly discern the high-frequency stripes in USAF target (Fig. 2a, green for 114, and red for 228 line pairs per mm). The GAN-microscopy reconstruction, in contrast, has easily resolved the finest part of USAF target (Fig. 2c2, 228 line pairs per mm). The GAN-reconstruction results are further compared with the realistic measurement under a  $\times 10$  magnification (Fig. 2b3, c3), showing a good structural similarity (SSIM) to the high-resolution ground truth. The linecuts through the resolved line pairs (Fig. 2c) by each method are quantified in Fig. 2d, revealing a substantially improved resolution by GAN. More quantitative analysis, such as SSIM and pixel signal to noise ratio (PSNR), of the GAN super-resolved image are described in Supplementary Table 1.



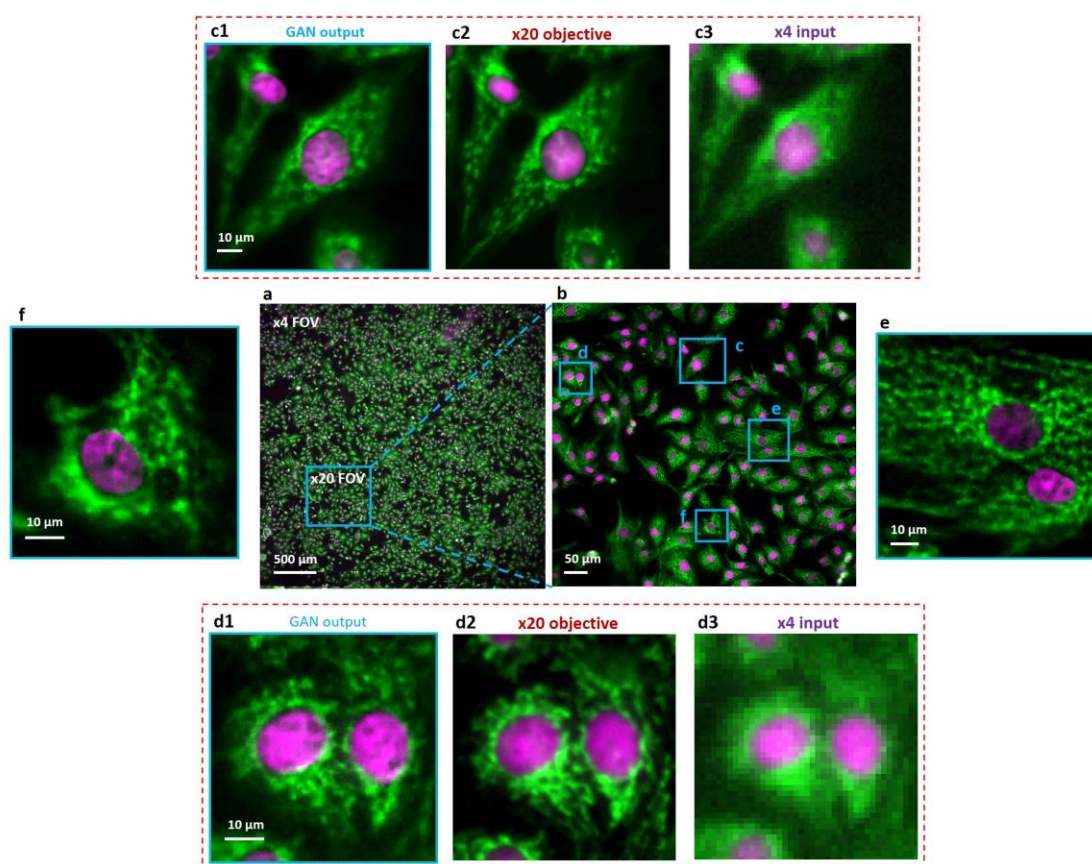
**Figure 2. Resolution characterization using USAF target.** **a**, raw image of the USAF resolution target taken under x2 low-magnification of a macro-zoom microscope. **b1** and **c1**, Magnified views of the raw image, with a pixel size of 2.13 μm. **b2** and **c2**, Reconstructions of a well-trained GAN-generator, with an enhancement factor of 5 (reconstructed pixel size, 0.425 μm). **b3** and **c3**, Corresponding high-resolution image taken under x10 magnification. **e**, Intensity plot of linecuts (shown in c1 and c2) for each method, indicating that GANM prototype provides substantively improved contrast and resolution (FWHM ~1.7 μm) that potentially enable subcellular-level imaging across a centimeter large scale.

## 1.2 Dual-color GAN-microscopy of fluorescent BPAE Cells

We demonstrate GAN-microscopy with our prototype by acquiring a wide-FOV dual-channel fluorescent image of a bovine pulmonary artery endothelial cell (BPAE) slide, as shown in Fig. 3a. The DAPI and Alexa Fluor 488 tagged nucleus and micro-filaments of the cells are 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 include an image degrading model in GAN network instead of the image registration. The model accurately characterizes the transfer function of microscope system (Olympus IX73), producing blurred and undersampled LR images in accordance with



their measured HR counterparts (Supplementary Note2). Practically, we generate 4-times lower-resolution simulations using  $\times 20$  realistic images, to constitute the  $\times 5$  and  $\times 20$  training data pairs with their FOV intrinsically aligned. Vignette high-resolution views of GAN-microscopy are provided in Fig. 3c-f with a reconstructed pixel size of  $0.325\ \mu\text{m}$ . The imaging FOV is around  $24\ \text{mm}^2$ , the same as that from a  $\times 4$  objective (Plan APO, 0.1 NA, Olympus), whereas the maximum achieved resolution is similar to that of a typical  $\times 20$  objective (Plan APO, 0.45 NA, Olympus). The conventional microscope images taken with  $\times 20$  and  $\times 4$  lenses are shown for comparison in Fig. 3c2, c3, d2, d3. Beside the superior SISR capability, it is noteworthy that GAN-reconstructed image also shows a large depth of field (DOF) which is succeeded from its  $\times 4$  source measurement. As a result, it behaves even better than the  $\times 20$  HR image in regions such as c2, where  $\times 20$  measurement is slightly out of focus. This underlying robustness of GAN-microscopy indicates its easy implementation on a broad range of samples with natively unsmooth surfaces.

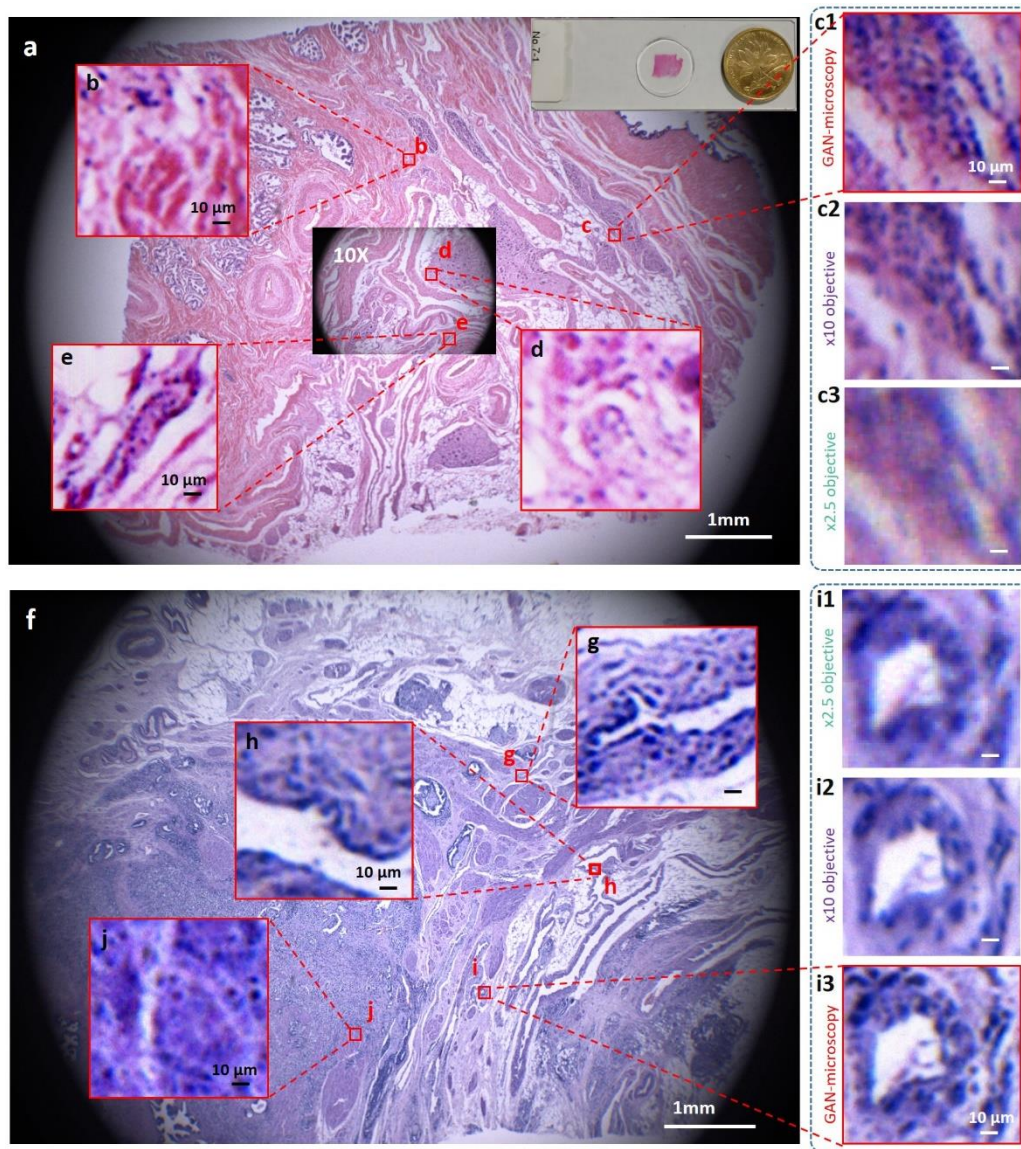


**Figure 3. Dual-color fluorescence imaging via GAN-microscopy.** **a**, A wide-FOV fluorescence image of BPAE cells specifically labelled with DAPI and Alexa Fluor-488, at nucleus and skeletons, respectively. **a** and **b** show the imaging FOV of  $\times 4$  (GAN-microscopy) and  $\times 20$  objectives (high-resolution conventional microscopy), by using a sCMOS camera (sensor area  $1.33 \times 1.33\ \text{cm}$ ). **c1**, **d1**, **e** and **f**, high-resolution views of the selected regions (blue) in **a**. **c1**, **d1** are further compared with **c2**, **c3**, **d2** and **d3**, which are taken under a conventional wide-field fluorescence microscope with a  $\times 20/0.45$  (**c2**, **d2**) and  $\times 4/0.1$  (**c3**, **d3**) objectives. High-resolution cells in **d2** experience slight out-of-focus due to the use of large NA objective, in contrast, rendering the advantage of GAN-microscopy (**d1**) that encompasses a large DOF.

### 1.3 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 GAN-microscopy to the imaging of human tissue sections that are from a normal prostate and a prostate tumor. HR reference images are taken by a macro-zoom microscope (Olympus mvx10), with using x10 magnification plus a color camera (QHY247C, 3.9  $\mu\text{m}$  pitch size). The corresponding x2.5 LR training images are down-sampled from the HR measurements via the degrading model. The real x2.5 measurements of the tissue slides have a large FOV of about 64  $\text{mm}^2$ , which almost covers the entire pathological sections, as shown in Fig. 4a, and f. On the other hand, the high-frequency details are decimated as shown in Fig. 4c3, i1. As comparison, the x10 HR images contain abundant details as illustrated in Fig. 4c2, i2, but suffer from much smaller FOV (4  $\text{mm}^2$ , inset in Fig. 4a). GAN-microscopy, however, is capable of providing numerous subcellular details across the entire tissue scale. The reconstructed pathology slides encompass 0.4 gigapixels, with an effective SBP being  $\sim 0.1$  gigapixels, which is 16-times higher than that of a conventional microscope. Vignette high-resolution views from different areas of the tissues are shown in Fig. b to e, and g to j, to verify this capability. 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.





**Figure 4. Gigapixel color imaging of prostate tissue slides.** **a** and **f**, GAN-microscopy (GANM) color images of normal prostate histology slide and prostate cancer pathology slide, respectively. The achieved effective SBP here is  $\sim 0.1$  gigapixels. **b-e**, **g-j** shows the vignette high-resolution views of the image in **a** and **f**. **c2-c3**, and **i2-i3**, Images taken by an Olympus macro-zoom microscope under x10 and x2.5 magnifications, for comparison with GAN results.

## 2. Discussion

We have demonstrated a deep learning-based microscopy method which can significantly improve the resolution of a conventional microscope and increase the imaging throughput for whole biomedical specimens. We apply a state-of-the-art GAN network to deeply learn a number of high-resolution microscopy images of prior examples, as well as their low-resolution counterparts. For cell images that contain complicated patterns, their low-resolution training data are artificially generated and automatically 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. After appropriate GAN training being accomplished, the well-established AI is capable of quickly estimating 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 the 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%. We also proved that this GAN-microscopy method is very robust, readily applicable to most forms of image data (bright-field imaging, fluorescence imaging, color imaging), significantly extending their space bandwidth product at least 16 folds, with neither at the cost of acquiring multiple frames nor relying on the retrofit of conventional microscope system. Therefore, GAN-microscopy has a high temporal performance similar to previous example-based SISR methods, 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  $\mu\text{m}$  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 GAN-microscopy 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 2D wide-field imaging of *exvivo* samples, we can reasonably expect that given its superior spatial-temporal performance, this methodology will be also applicable to both 3-D microscopy and highly dynamic process.

### 3. Methods

#### 3.1 Generated Adversarial Network Model

The single image super resolution aims to estimate a high-resolution, super-resolved image  $I^{SR} \in \mathbb{R}^{rW \times rH \times C}$  given a low-resolution image  $I^{LR} \in \mathbb{R}^{W \times H \times C}$ , where  $W$ ,  $H$  and  $C$  are the width, height and color channels of the image, respectively, and  $r$  is the up-scaling factor. Here  $I^{LR}$  is the low-resolution version of its high-resolution counterpart  $I^{HR}$ . Our ultimate goal is to make the estimated super-resolved image  $I^{SR}$  be perceptually as similar to the original high-resolution version  $I^{HR}$  as possible.

In order to achieve this goal, we train a generator network  $G_{\theta_G}$  that takes as input the raw low-resolution image and outputs the corresponding high-resolution counterpart. This generator network is based on the feed-forward convolution neural network (CNN) and parameterized by  $\theta_G$ . To obtain the optimal parameters  $\theta_G$ , previous most widely used method is to minimize the pixel-wise error measurements such as the mean squared error (MSE) between the estimated image and the ground truth. However, while this strategy is able to achieve high peak signal-to-noise ratio (PSNR), the solutions of MSE based optimization look perceptually unsatisfactory with high-frequency information missed and over-smoothed textures occupied. In order to alleviate this problem, we followed the generative adversarial network (GAN) work from Goodfellow et al<sup>31</sup>. by further incorporating a discriminator network  $D_{\theta_D}$ . We optimize the two networks  $G_{\theta_G}$  and  $D_{\theta_D}$  simultaneously in an alternating manner by solving the adversarial min-max problem :

$$\min_{\theta_G} \max_{\theta_D} E_{I^{HR} \sim P_{train}(I^{HR})} [\log D_{\theta_D}(I^{HR})] + E_{I^{LR} \sim P_G(I^{LR})} [\log (1 - D_{\theta_D}(G_{\theta_G}(I^{LR})))]$$

where  $D_{\theta_D}(I)$  is the probability that any given image  $I$  is classified into a natural high-resolution image by the discriminator  $D$ . The basic idea of this GAN formula is that the generator model  $G$  is trained to fool the discriminator  $D$  that is trained to differentiate between the generated images by  $G$  and the real images. In this way, the solutions of our generator  $G$  can highly resemble real images so that it is hard for the discriminator to tell apart. We believe this method can lead to perceptually superior solutions that reside in the manifold of natural images. The architecture and loss function of the generator and discriminator will be discussed in details in the first section of Supplementary Information.

### 3.2 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 (Thermofisher scientific), 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.

### 3.3 Training dataset generation

It is widely accepted that the performance of a neural network relies heavily on the training dataset, where there are LR images as input data and HR images as target labels for super resolution task. These LR and HR image pairs for training can be obtained in two ways. The first and the most intuitive method is that both LR and HR images can be experimentally captured with the microscope. However, since the LR and HR image pairs are taken under different microscope 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 and tissue imaging, a great deal of feature details are 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 dataset.

On the other hand, 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. In this case, the captured LR images are only used in the model validation phrase. 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. Specifically,

it consists of optical blurring, down-sampling and statistical noise addition operations, which simulates the realistic decimation process from HR to LR in imaging. The down-sampling operation is performed via bicubic interpolation with a factor of  $1/n$ , where  $n$  is the magnification between LR and HR. For image blurring and noise addition, we used the following practices: the blurring is implemented by convoluting the image with a Gaussian filter (also known as the point spread function) and the noise we added is Gaussian white noise with mean value of 0. There are two other parameters to be determined by the actual optical imaging system: the size of the Gaussian kernel and the variance of the white noise distribution. We made trials and errors to figure out these two parameters for our system, which were summarized in the Section 2 of Supplementary Information.

After generating the simulated LR images using the captured HR measurements, we cut the images into small pieces to compose the training dataset, of which the LR pieces are  $96 \times 96$  pixels and HR pieces  $384 \times 384$  pixels. Each LR piece has a corresponding HR one with the same FOV. By cutting the images, we can increase the number of training data by a lot, and at the same time, smaller image size is more compatible to the limited GPU memory. We also adopted several widely-used data augmentation methods, such as translation, rotation and flipping, to further expand the scale of training dataset.

### 3.4 Training implementations

Our model was implemented based on the Google's deep learning framework, TensorFlow (version r1.3), and trained on an Inspur® server with two NVidia Tesla P100 GPUs. Initiating with a batch size of 16 and a fixed learning rate of  $1e-4$ , we trained the network for 200 epochs, which took about 48 hours.

### 3.5 Inference process

In the inference phase after network training, the experimentally captured LR images for validation were 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 were 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  $100 \times 100$  pixels size takes less than 0.01 second to be super-resolved into a  $400 \times 400$  pixels image, even on an ordinary Windows laptop with Intel Core i5 CPU.

### 3.6 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, and the bright-field color images of two types of tissue slides. Images of resolution target were recorded by a Photometrics IRIS15 camera (pitch size  $4.25 \mu\text{m}$ ), with the HR and LR images taken under  $\times 10$  and  $\times 2$  magnifications of an Olympus MVX10 microscope, respectively. The BPAE cells were specifically labeled with Alexa Fluor-488 and DAPI, at skeletons and nucleus, before they were

imaged under an Olympus IX73 microscope equipped with a HAMAMATSU ORCA-Flash 4.0-V2 camera (pitch size 4.25  $\mu\text{m}$ ). In both fluorescent channels, the HR training images and LR validation images were taken under a x20/0.45 and x4/0.1 objective, respectively. For pathology slide imaging, a QHY247C color camera (pitch size 3.9  $\mu\text{m}$ ) were used on the Olympus MVX 10 microscope to capture the healthy prostate/prostate cancer tissues that are stained with hematoxylin-eosin. The HR training and LR validation images were then taken under x10 and x2 magnifications, respectively.

## Acknowledgements

The authors acknowledge the selfless sharing of the GAN source codes from Hao Dong (hao.dong11@imperial.ac.uk), as well as the contributions of Tinting Zhu for assistance with fluorescent sample preparation. The authors thank Wenbin Jiang and Yang Ma for their assistance with GPU-based computation. This research has received funding support from 1000 Youth Talents Plan of China (P.F.), Fundamental Research Program of Shenzhen (P.F., JCYJ20160429182424047)

## Author Contributions

P.F. and D.J. conceived the idea, initiated the investigation. H.Z., X.X., C.F. and Y.Y. developed the programs, acquired and processed data. H.Z., D.J. and P.F. prepared the manuscript.

## Additional Information

Correspondence and requests for materials should be addressed to feipeng@hust.edu.cn

## Competing Financial Interests

The authors declare no competing financial interests.

# Supplementary Information

## 1. SRGAN architecture

### 1.1 Generator

The model architecture of the generator  $G$  is illustrated in Supplementary Figure 1(a). In details, the input image is first processed by one convolution layer of  $3 \times 3$  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  $3 \times 3$  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.<sup>32</sup>.

To optimize this model, we designed the perceptual loss function that is the weighted sum of the MSE loss  $\ell_{MSE}^G$ , the feature reconstruction loss  $\ell_{feat}^G$  and the adversarial loss  $\ell_{adv}^G$  as proposed by Ledig et al.<sup>28</sup>:

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

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:

$$\ell_{MSE}^G(G_{\theta_G}(I^{LR}), I^{HR}) = \frac{1}{r^2 WH} \sum_{x=1}^{rW} \sum_{y=1}^{rH} (I_{x,y}^{HR} - G_{\theta_G}(I^{LR})_{x,y})^2.$$

Besides using the GAN framework to encourage perceptual similarity, we further used the special feature reconstruction loss function proposed by Johnson et al.<sup>33</sup>. Let  $\phi_j(x) \in R^{W_j \times H_j \times C_j}$  be the activations of the  $j$ th convolution layer of the VGG19 network described in Simonyan and Zisserman<sup>34</sup> when processing the image  $x$ . Then the feature reconstruction loss is defined by the Euclidean distance between the feature representations of the reconstructed image  $G_{\theta_G}(I^{LR})$  and the reference image  $I^{HR}$ :

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

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:

$$\ell_{adv}^G(G_{\theta_G}(I^{LR})) = -\log D_{\theta_D}(G_{\theta_G}(I^{LR})),$$

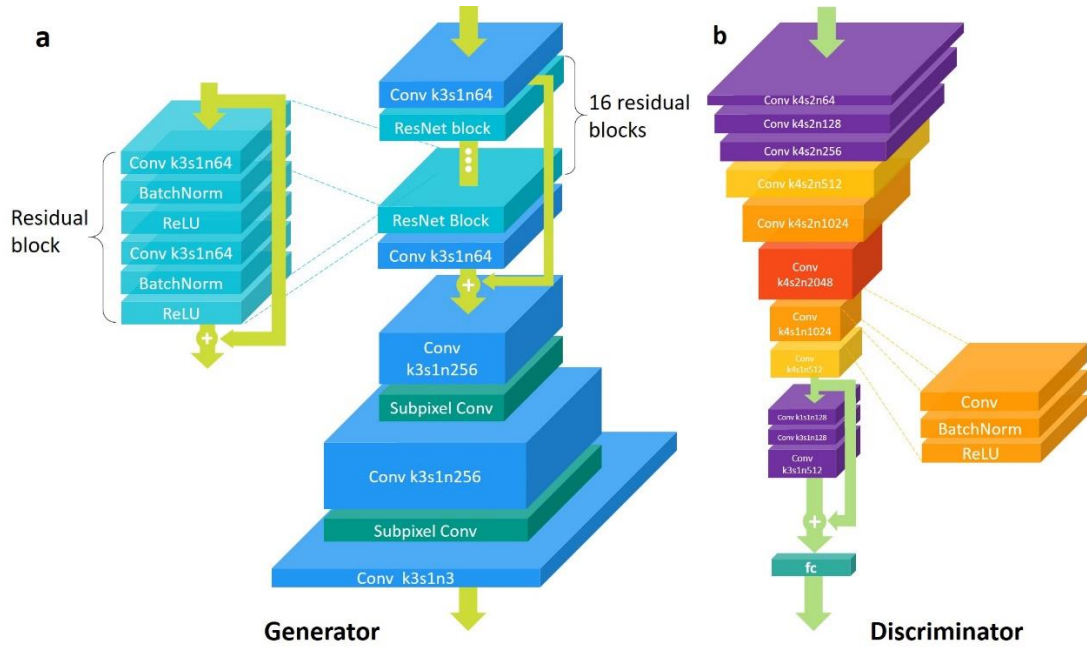
where for better gradient computation stability<sup>31</sup>, we minimize  $-\log D_{\theta_D}(G_{\theta_G}(I^{LR}))$  instead of  $\log(1 - D_{\theta_D}(G_{\theta_G}(I^{LR})))$ .

## 1.2 Discriminator

The architecture of the discriminator  $D$  is shown in Supplementary Figure 1(b). It first contains 8 convolutional layers with  $4*4$  kernels followed by BN layers and LeakyReLU ( $\alpha=0.2$ ) 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:

$$\ell^D = \frac{1}{N} \sum_{n=1}^N (-\log D_{\theta_D}(I_n^{HR}) - \log(1 - D_{\theta_D}(G_{\theta_G}(I_n^{LR}))))$$

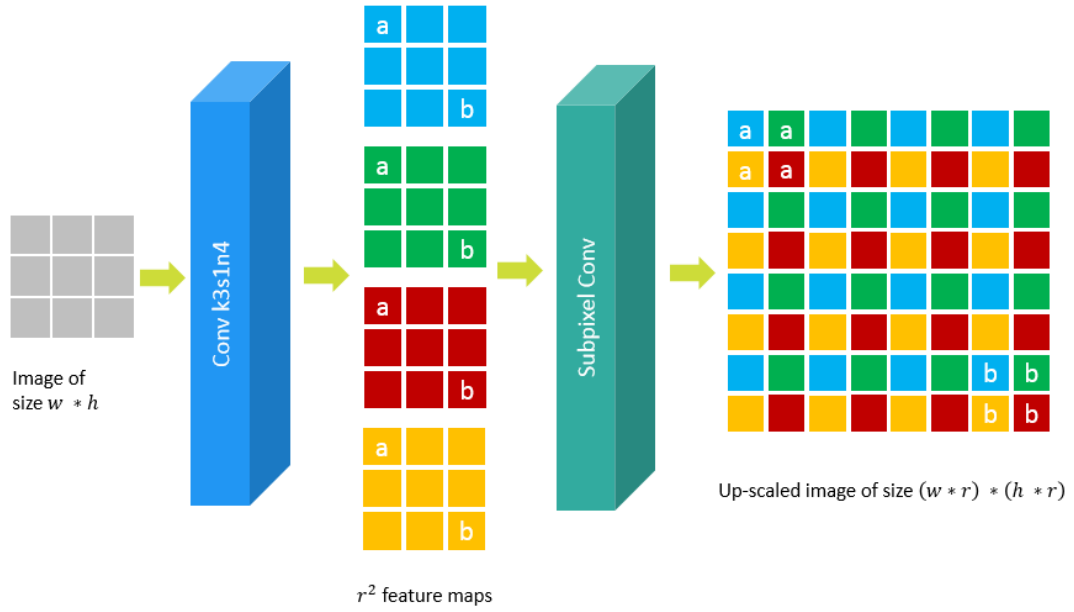




**Supplementary Figure 1. The architecture of GAN.** **a**, Architecture of the generator. Conv and ResNet is the abbreviation of Convolutional layer and Residual network block. The parameters of the convolutional layers is given in the format " $k$ - $s$ - $n$ ", where  $k$  is the kernel size,  $s$  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**, -12-layer 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.

### 1.3 Interpolation LR inputs

Unlike a beforehand interpolation outside the network<sup>35</sup> or the use of a deconvolution layer<sup>36</sup>, GAN applies a novel architecture called sub-pixel convolution layer to generate up-scaled version of the input image inside the network. Its implementation is shown in following Supplementary Figure 2. In order to illustrate how a subpixel convolution layer work, we simplify the input layer to a single-channel image of  $w$  by  $h$  pixels. The input is first extracted to multiple channels by a convolutional layer. Based on variant kernels, the convolution layer with  $r^2$  feature maps generate  $r^2$  channels of the input. Each channel has the same width and height as the input LR image. The Subpixel convolution layer then fills the pixels from all these channels into a one-channel, high-resolution grid, according to their spatial correlations. Finally, an interpolation image with dimension being  $(w*r)$  by  $(h*r)$  is generated as the output of subpixel convolution layer. Along with this logic, if an image with  $m$  channels is to be up-scaled by a factor  $r$ , there must be  $m * r^2$  channels in the direct input layer of the subpixel convolution layer. The pixels in these channels are rearranged in a high resolution grid in their original order, fusing into an  $m$ -channel, interpolated output.



**Supplementary Figure 2. Structure of subpixel convolution layer.** The subpixel convolution layer together with a convolution layer accurately transform the input images into an interpolated ones, by a factor of  $r$ .

## 2. The degrading model

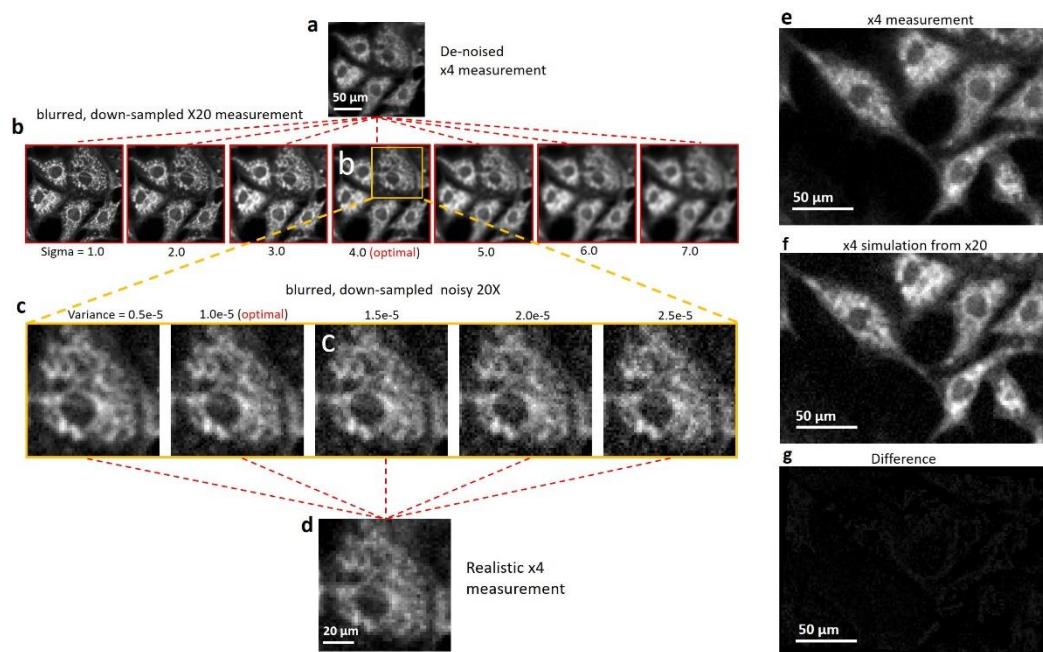
To generate simulated microscopy data that fairly close to the realistic measurements, the parameters in the degrading model need to be finely tuned. The degrading process of conventional microscopy system can be described as

$$I_m = D(K * I) + N$$

of which  $I$  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 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 need to be optimized: size of the Gaussian kernel in the convolution step and variance of the noise distribution. Taking the degradation of a x20 cell image for example, we first de-noise the x4 measurement, comparing it with the blurred then down-sampled x20 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 x4 measurement to identify the proper variance. After the model being created, we verify it multiple times. The difference between

the simulation and the measurement are directly computed as the pixel-wise subtraction, as shown in Fig. S3g. 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.



**Supplementary Figure 3. Generation of simulated LR inputs via a degradation model.** **a**, De-noised x4 LR measurement used for finding the optimized blurring kernel of microscope. **b**, The blurred and down-sampled images from x20 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 x4 measurement. The best matched level of noise is found as a result. **e to g**, a x4 measurement is subtracted by the optimized x4 simulation to verify the efficacy of the model.

### 3. Quantitative evaluation of image similarity

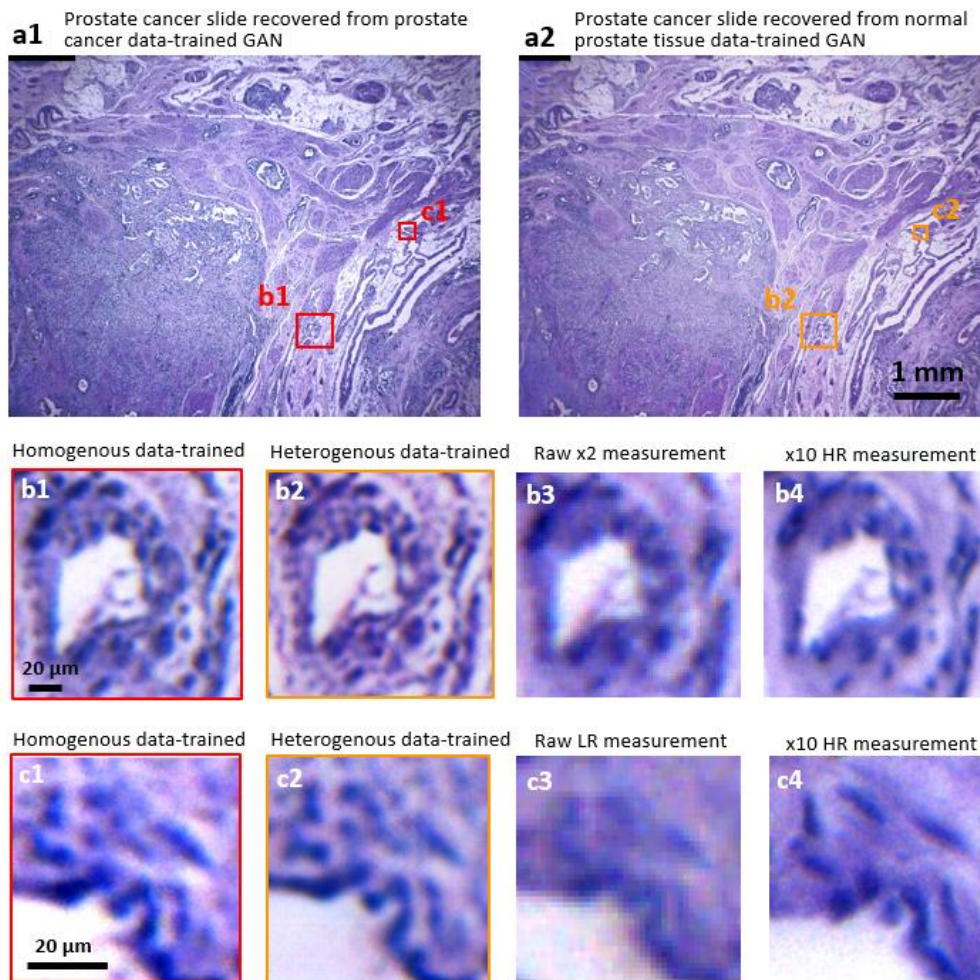
We calculate the peak signal-to-noise ratio (PSNR[dB]) and the structural similarity index (SSIM)<sup>37</sup> between the GAN-microscopy outputs and the high-resolution measurements (Supplementary Table 1), to quantify the accuracy of the GAN-microscopy. For comparison, PSNR and SSIM index between the bicubic interpolation 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  $\infty$  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 values of all the 16 as final results. Apparently, SSIM and PSNR of the network reconstructions are both far better than that of the 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.

	SSIM	PSNR
Bicubic Interpolation	0.8093	19.6299
Network Reconstruction	0.9317	27.2610
High Resolution Target	1	$\infty$

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

#### 4. The robustness of the network

The GAN network training is fundamentally a process in which the neural network extracts features from the training data and learns their roles in the corresponding low-resolution and high-resolution 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 the SR estimation of 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 Supplementary Figure 4 below. It is remarkable that both networks recover highly similar structures with similar qualities presented. Both networks are 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 GAN-microscopy can go further with being applied to the reconstruction of a variety of samples merely with single type of data training.



**Supplementary Figure 3. Validating the robustness of GAN-microscopy.** **a**, Wide-FOV GAN images of a prostate cancer slide using homogenous-data-trained (prostate cancer images) network and Heterogenous-data-trained (healthy prostate images) network. **b1-c1**, and **b2-c2**, Vignette high-resolution views of the image in **a1** and **a2**, respectively. **b3-c3**, and **b4-c4**, Images taken by a macro-zoom microscope (mvx10) under a x2.5 and a x10 magnification, respectively, for comparison.

## Reference

- 1 Zheng, G., Horstmeyer, R. & Yang, C. Wide-field, high-resolution Fourier ptychographic microscopy. *Nature Photonics* **7**, 739 (2013).
- 2 Lohmann, A. W., Dorsch, R. G., Mendlovic, D., Ferreira, C. & Zalevsky, Z. Space-bandwidth product of optical signals and systems. *Journal of the Optical Society of America A* **13**, 470-473 (1996).

- 3 Imfeld, K. *et al.* Large-Scale, High-Resolution Data Acquisition System for Extracellular Recording of Electrophysiological Activity. *IEEE Transactions on Biomedical Engineering* **55**, 2064 (2008).
- 4 Brown, M. & Lowe, D. G. Automatic Panoramic Image Stitching using Invariant Features. *International Journal of Computer Vision* **74**, 59-73 (2007).
- 5 Hillman, T. R., Gutzler, T., Alexandrov, S. A. & Sampson, D. D. High-resolution, wide-field object reconstruction with synthetic aperture Fourier holographic optical microscopy. *Optics Express* **17**, 7873 (2009).
- 6 M, K. *et al.* High-speed synthetic aperture microscopy for live cell imaging. *Optics letters* **36**, 148 (2011).
- 7 Gutzler, T., Hillman, T. R., Alexandrov, S. A. & Sampson, D. D. Coherent aperture-synthesis, wide-field, high-resolution holographic microscopy of biological tissue. *Optics Letters* **35**, 1136 (2010).
- 8 Luo, W., Greenbaum, A., Zhang, Y. & Ozcan, A. Synthetic aperture-based on-chip microscopy. *Light Science & Applications* **4** (2015).
- 9 Zheng, G., Lee, S. A., Yang, S. & Yang, C. Sub-pixel resolving optofluidic microscope for on-chip cell imaging. *Lab on A Chip* **10**, 3125-3129 (2010).
- 10 Zheng, G., Lee, S. A., Antebi, Y., Elowitz, M. B. & Yang, C. The ePetri dish, an on-chip cell imaging platform based on subpixel perspective sweeping microscopy (SPSM). *Proceedings of the National Academy of Sciences of the United States of America* **108**, 16889 (2011).
- 11 Luo, W., Zhang, Y., Feizi, A., Göröcs, Z. & Ozcan, A. Pixel super-resolution using wavelength scanning. *Light Science & Applications* **5**, e16060 (2016).
- 12 Xu, W., Jericho, M. H., Meinertzhagen, I. A. & Kreuzer, H. J. Digital in-line holography for biological applications. *Proceedings of the National Academy of Sciences of the United States of America* **98**, 11301-11305 (2001).
- 13 Denis, L., Lorenz, D., Thiébaud, E., Fournier, C. & Trede, D. Inline hologram reconstruction with sparsity constraints. *Optics Letters* **34**, 3475 (2009).
- 14 Greenbaum, A. *et al.* Increased space-bandwidth product in pixel super-resolved lensfree on-chip microscopy. *Scientific Reports* **3**, 1717 (2013).
- 15 Elad, M. & Hel-Or, Y. A fast super-resolution reconstruction algorithm for pure translational motion and common space-invariant blur. *IEEE transactions on image processing : a publication of the IEEE Signal Processing Society* **10**, 1187 (2001).
- 16 Farsiu, S., Robinson, M. D., Elad, M. & Milanfar, P. Fast and robust multiframe super resolution. *IEEE Transactions on Image Processing* **13**, 1327-1344 (2004).
- 17 Vandewalle, P., Sü, S. & Vetterli, M. A Frequency Domain Approach to Registration of Aliased Images with Application to Super-Resolution. *Eurasip Journal on Advances in Signal Processing* **2006**, 1-14 (2006).
- 18 Gustafsson, M. G. L. Surpassing the lateral resolution limit by a factor of two using structured illumination microscopy. *Journal of Microscopy* **198**, 82-87 (2010).
- 19 MG, G. Nonlinear structured-illumination microscopy: wide-field fluorescence imaging with theoretically unlimited resolution. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 13081 (2005).



- 20 Yang, J., Wright, J., Huang, T. & Ma, Y. Image super-resolution as sparse representation of raw image patches. *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on.* 1-8.
- 21 Timofte, R., Smet, V. D. & Gool, L. V. A+: *Adjusted Anchored Neighborhood Regression for Fast Super-Resolution.* (Springer International Publishing, 2014).
- 22 Cireşan, D. C., Giusti, A., Gambardella, L. M. & Schmidhuber, J. Mitosis detection in breast cancer histology images with deep neural networks. 411-418.
- 23 Dan, C. C., Giusti, A., Gambardella, L. M. & Schmidhuber, J. Deep Neural Networks Segment Neuronal Membranes in Electron Microscopy Images. *Advances in Neural Information Processing Systems* **25**, 2852--2860 (2012).
- 24 Cruzroa, A. *et al.* 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).
- 25 Litjens, G. *et al.* Deep learning as a tool for increased accuracy and efficiency of histopathological diagnosis. *Scientific Reports* **6**, 26286 (2016).
- 26 Ertosun, M. G. & Rubin, D. L. Automated Grading of Gliomas using Deep Learning in Digital Pathology Images: A modular approach with ensemble of convolutional neural networks. *AMIA ... Annual Symposium proceedings. AMIA Symposium* **2015**, 1899 (2015).
- 27 Xu, J., Luo, X., Wang, G., Gilmore, H. & Madabhushi, A. A Deep Convolutional Neural Network for segmenting and classifying epithelial and stromal regions in histopathological images. *Neurocomputing* **191**, 214 (2016).
- 28 Ledig, C. *et al.* Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network. (2016).
- 29 Lim, B., Son, S., Kim, H., Nah, S. & Lee, K. M. Enhanced deep residual networks for single image super-resolution. *Computer Vision and Pattern Recognition Workshops.* 1132-1140.
- 30 Dahl, R., Norouzi, M. & Shlens, J. Pixel Recursive Super Resolution. (2017).
- 31 Goodfellow, I. J. *et al.* Generative adversarial nets. *International Conference on Neural Information Processing Systems.* 2672-2680.
- 32 Shi, W. *et al.* Real-Time Single Image and Video Super-Resolution Using an Efficient Sub-Pixel Convolutional Neural Network. 1874-1883 (2016).
- 33 Johnson, J., Alahi, A. & Fei-Fei, L. Perceptual Losses for Real-Time Style Transfer and Super-Resolution. 694-711 (2016).
- 34 Simonyan, K. & Zisserman, A. Very Deep Convolutional Networks for Large-Scale Image Recognition. *Computer Science* (2014).
- 35 Dong, C., Chen, C. L., He, K. & Tang, X. Image Super-Resolution Using Deep Convolutional Networks. *IEEE Transactions on Pattern Analysis & Machine Intelligence* **38**, 295-307 (2014).
- 36 Zeiler, M. D., Taylor, G. W. & Fergus, R. Adaptive deconvolutional networks for mid and high level feature learning. *International Conference on Computer Vision.* 2018-2025.
- 37 Wang Z., B. A. C., Sheikh H. R., Simoncelli E. P. . Image Quality Assessment: From Error Visibility to Structural Similarity. *IEEE Trans. Image Process* **13**, 13 (2004).