bioRxiv preprint doi: https://doi.org/10.1101/848077; this version posted November 21, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Unsupervised content-preserving transformation for optical microscopy

Xinyang Li[1,2,3†], Guoxun Zhang[1,2†], Jiamin Wu[1,2], Hao Xie[1,2], Xing Lin[1,2,4], Haoqian Wang[2,3], Hui Qiao[1,2*] & Qionghai Dai[1,2*]

[1]*Department of Automation, Tsinghua University, Beijing 100084, China*

[2]*Institute for Brain and Cognitive Sciences, Tsinghua University, Beijing 100084, China.*

[3]*Tsinghua Shenzhen International Graduate School, Tsinghua University, Shenzhen, 518055, China*

[4]*Beijing Innovation Center for Future Chip, Tsinghua University, Beijing 100084, China*

†*These authors contributed equally to this work.*

*Correspondence: qhdai@mail.tsinghua.edu.cn, qiaohui@mail.tsinghua.edu.cn

bioRxiv preprint doi: https://doi.org/10.1101/848077; this version posted November 21, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

The advent of deep learning and the open access to a substantial collection of imaging data provide a potential solution to computational image transformation, which is gradually changing the landscape of optical imaging and biomedical research. However, current deep-learning implementations usually operate in a supervised manner, and the reliance on a laborious and error-prone data annotation procedure remains a barrier towards more general applicability. Here, we propose an unsupervised image transformation enlightened by cycle-consistent generative adversarial networks (cycleGANs) to facilitate the utilization of deep learning in optical microscopy. By incorporating the saliency constraint into cycleGAN, the unsupervised approach, dubbed as content-preserving cycleGAN (c^2GAN), can learn the mapping between two image domains and avoid the misalignment of salient objects without paired training data. We demonstrate several image transformation tasks such as fluorescence image restoration, whole-slide histological coloration, and virtual fluorescent labeling. Quantitative evaluations prove that c^2GAN achieves robust and high-fidelity image transformation across different imaging modalities and various data configurations. We anticipate that our framework will encourage a paradigm shift in training neural networks and democratize deep learning algorithms for optical society.

Deep learning[1] has made great progress in computational imaging and image interpretation[2,3]. As a data-driven methodology, deep neural networks with high model capacity can theoretically approximate arbitrary function that represents the mapping from the input domain to the output domai[4,5]. Given images as the inputs in high-dimensional space, the applications of deep learning in optical microscopy can be divided into two categories according to the forms of the outputs.

bioRxiv preprint doi: https://doi.org/10.1101/848077; this version posted November 21, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

The first one is image classification[6,7] where deep neural networks are trained to classify an image in terms of its visual content. Such applications include cell sorting[8,9], morphology recognition[10,11], state identification[12] and pathological diagnosis[13,14]. The other one is image transformation whose outputs are still images but with highlighted or previously inaccessible information, aiming to visualize imperceptible structures and latent patterns, as well as expand the design space of common imaging systems[15].

Recently, several network architectures have been adopted for image transformation to learn the mapping between a source domain and a target domain. U-Net[16] is one of the most popular convolutional neural networks (CNNs) for pixel-wise regression, which is demonstrated to have the state-of-the-art performance in cell segmentation and detection[17], image restoration[15], and 3D virtual fluorescence labeling[18]. Some elaborately modified CNNs can also estimate the mapping functions and complete the tasks of image transformation like resolution improvement[19] and virtual labeling[20]. Moreover, generative adversarial network (GAN), an emerging deep learning framework based on minimax game theory that trains a generative model and an adversarial discriminative model simultaneously[21,22], can learn to optimize a perceptual-level loss function and produce more realistic results. GANs have been verified feasible for different transformations, such as super-resolution reconstruction[23,24], bright-field holography[25], and virtual histological staining[26].

Regarding current deep learning algorithms, the superior performance and stability on the test set largely depend on a high-quality collection of substantial microscopy images, which is the most crucial factor for the success of the training process. In the conventional strategy of supervised learning, vast amounts of images and corresponding annotations are necessary, which is time-consuming and error-prone when performed manually, especially for image

bioRxiv preprint doi: https://doi.org/10.1101/848077; this version posted November 21, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

transformation tasks with the need of pixel-level alignment. Although data augmentation and transfer learning have been widely employed to reduce the size of the training set, collecting a small number of aligned image pairs still necessitates massive modifications to the imaging systems and some complicated experimental procedures. In some cases, strictly aligned training pairs are impossible to obtain because of the significant differences between imaging mechanisms. In essence, the dilemma between the indispensability of paired training data and the dearth of annotated datasets obstructs the advancement of deep learning in optical microscopy. Nowadays, the invention of cycle-consistent generative adversarial networks (cycleGANs) makes unsupervised training of CNNs possible[27]. CycleGAN uses a cycle-consistency learning technique to translate images from one domain to the other one and exhibits comparable performance to supervised methods. Such learning strategy and its modified versions have been validated in style transfer of natural images[28-30] and medical image analysis[31-33]. As for the field of optical microscopy, a few forward-looking studies have applied cycleGANs to remove coherent noise in optical diffraction tomography[34], and realize image segmentation for bright-field microscopy and X-ray computed tomography[35].

To advance the feasibility of this unsupervised learning framework in optical microscopy, we propose a content-preserving cycleGAN (c^2GAN) for precise image-to-image transformation. At a conceptual level, the principle of c^2GAN is depicted in Fig. 1a. Along with the novel cycle consistency loss[27], an additional saliency constraint is enforced to locate the salient objects (*i.e.*, the image contents) because the cycle consistency loss alone is not rigorous enough to guarantee satisfactory solutions (see Methods). The saliency constraint can reshape the spatial manifold of the objective function and shrink the solution space to a tighter range to effectively exclude

bioRxiv preprint doi: https://doi.org/10.1101/848077; this version posted November 21, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

unwanted solutions that are likely to result in biased mappings and producing unendurable transformed images with risks of misguiding biological analyses (see Fig. 1b).

The workflow of c^2GAN is schematized in Fig. 1c. Like most machine learning methodologies, the whole pipeline includes the training phase and the prediction phase. In the training phases, two image sets (*A* and *B*) are first collected to sample two image domains of distinct modalities (*e.g.* low-SNR and high-SNR, grayscale and RGB, bright-field and fluorescent, *etc.*). No pre-aligned image pairs are required in these two image collections. Then, a forward GAN is trained to transfer images in *A* into the style of *B* and a forward adversarial loss is used to evaluate the quality of transformation. Similarly, a backward GAN is simultaneously trained to transfer images in *B* into the style of *A* and a backward adversarial loss is used to evaluate the quality of transformation. The cycle consistency term is used to guide the twin GANs to form a closed cycle. Additionally, a saliency constraint is designed to impose restrictions on network outputs to correct some possible mapping bias. The whole training process works in an unsupervised learning manner. After proper training, an image could be mapped back through the sequential processing of the two GANs, which indicates the success of training. In the prediction phase, the generative model of the pre-trained forward GAN is loaded and some data never seen by the network will be fed into the model to check its generalization ability.

bioRxiv preprint doi: https://doi.org/10.1101/848077; this version posted November 21, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Results

We validated c^2GAN on various image transformation tasks based on some publicly available data. These tasks could be regarded as learning to perform nonlinear pixel-wise regression between two spaces of different dimensions. Although c^2GAN completes the task of unaligned domain mapping, ground truth is still needed in the test phase to evaluate the results and make convincing interpretations. In this section, we demonstrate our experimental results on fluorescence image restoration (including denoising, axial resolution restoration, and super-resolution reconstruction), whole-slide histological coloration, and virtual cell staining. We compared our results with corresponding ground truth, as well as the results of conventional supervised learning methods, to indicate the capability and reliability of c^2GAN.

Fluorescence image restoration. Firstly, we applied c^2GAN to the restoration of fluorescence images based on the experimental data released by Weigert *et al*[15]. In this series of tasks, both the input and the output images are single-channel greyscale images. The regression needed to learn is a mapping between two spaces of the same dimension. Considering the mechanism that there are two reciprocal GANs in c^2GAN, training the forward GAN and training the backward GAN thus have comparable complexity theoretically. For the denoising of confocal images of Planaria, we first randomly cropped half of the low-SNR images into around 20000 small patches of 128×128 pixels. Then did the same to the other half of high-SNR images. This step generated two image collections without any aligned image pairs. We trained c^2GAN with the two image collections to learn the mapping function between the domain of low-SNR and the domain of high-SNR. Then, some low-SNR images that never seen by the network were used to evaluate the model and the results are shown in Fig. 2a. Large images were split into small patches with 25%

bioRxiv preprint doi: https://doi.org/10.1101/848077; this version posted November 21, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

overlaps and the results were stitched back together. The stitched result shows that c^2GAN succeeds to learn the mapping function from the low-SNR domain to the high-SNR domain and the cell structures are well maintained at the same time. We compared the results of c^2GAN with that of the conventional supervised learning method (Supplementary Fig. 1). Because GANs are designed to generate certain distributions, the results of c^2GAN have higher fidelity and no pixel saturation occurs. Moreover, we reconstructed a piece of a surface from 2856 (7×8×51) image patches in Fig. 2b to show the overall effect of c^2GAN on volumetric data. The enhancement after denoising is remarkable and the whole data volume keeps a high consistency with the ground truth, which indicates the good generalization of our model.

Axial resolution degradation of most scanning microscopy[36] can also be restored by the proposed unsupervised learning framework. We trained c^2GAN with two totally unaligned image collections generated by the same procedure as denoising. As shown in Fig. 2c, an accurate mapping function for axial resolution restoration can be established after proper training. For a more quantitative analysis, we calculated the peak signal-to-noise ratio (PSNR) of all 15 test images, as well as the PSNR of the input images, shown in Fig. 2d. On average, image PSNR is improved by 5dB with the reconstruction of c^2GAN. We also verified our model on the restoration of multicolor zebrafish retina images with a different degradation coefficient to exhibit model capability (Supplementary Fig. 2).

Moreover, c^2GAN can also be used in resolving sub-diffraction structures such as microtubules and secretory granules (see Fig. 2e). A mapping function between the domain of wide-field microscopy and the domain of super-resolution microscopy can be learned without the supervision of any aligned training pairs. The results of c^2GAN still keep a high accuracy and fidelity compared with the CARE net (Supplementary Fig. 3). The intensity profiles of some fine

bioRxiv preprint doi: https://doi.org/10.1101/848077; this version posted November 21, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

structures in Fig. 2e are plot in Fig. 2f, which demonstrate that some structures beyond identification according to Rayleigh criterion in the raw images can be recognized again with the reconstruction of c^2GAN.

Whole-slide histological coloration. Next, we validated c^2GAN on image transformation from low-dimensional space to high-dimensional space. A typical task is image coloration[37,38] that maps a single-channel image to a three-channel image (*i.e.* black-and-white to RGB). For improving resolution and quantum efficiency, Bayer pattern is abandoned in most mainstream scientific cameras. Also, no matter for wide-field detection with cameras[39-41] or single-point detection with photomultiplier tubes (PMT)[42-44], multicolor imaging usually requires multi-channel detection with parallel-configured several detectors[45-46], which is complex and expensive. Image coloration explores the possibility of recovering spectral information from a single intensity image. This problem is considered as badly underconstrained and usually needs strong structural priors to solve it.

Here in c^2GAN, we model it as a pixel-wise regression problem that maps a single-pixel intensity value to a ternary vector with the assistance of neighborhood information. To make two unpaired image collections for unsupervised training, we downloaded whole-slide histopathology images of hematoxylin and eosin (H&E)-stained brain tissues from TCGA obtained by surgical excision followed by frozen section preparation. These patients were diagnosed with glioma of different grades. To avoid learning improper color mapping functions caused by divergent staining effects, we unified the color style of the dataset and finally, 26 whole-slide images (Supplementary Table 1) with the same color scheme were selected (Supplementary Fig. 4). To make the training set, we first divided 20 pathological slides into two equal parts (10 for domain *A* and 10 for domain *B*). Those images for domain *A* were converted to their black-and-white

bioRxiv preprint doi: https://doi.org/10.1101/848077; this version posted November 21, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

(BW) version. We randomly cropped both the RGB images, as well as the BW images into about 45000 patches with the size of 256×256 for each domain. This procedure could prevent producing aligned image pairs that are likely to mislead unsupervised training.

After training, another six whole-slide histopathology images that never seen by the network were used to test c^2GAN's performance (Supplementary Fig. 5). Five of them are also from glioma patients while the other one is from bronchus and lung tissue. These whole-slide images (the typical size is about 26000×22000 pixels) were first split into thousands of 256×256 patches with an overlap of 64 pixels, successively fed into the pre-trained network and then stitched again to form a whole-slide image. To further demonstrate the generalization ability of the pre-trained model, a slide instance of lung tissue slide that extends over a 1.3 cm × 1.1 cm area is shown in Figs. 3a-3c, which demonstrates the overall effect of whole-slide coloration. We also enlarge a small region with a high cell density in Figs. 3d-3f, where the nuclear and cytoplasmic distribution remains intact after coloration by c^2GAN. For error evaluation, we visualized the errors of each channel (Supplementary Fig. 6) and calculated corresponding structural similarity index (SSIM) with reference to corresponding ground truth. The quantitative evaluation indicates that c^2GAN can complete whole-slide coloration with SSIM as high as 0.99, which means that the proposed unsupervised learning framework can faithfully recover not only the overall color scheme but also the detailed color assignment without disturbing cell morphology.

Virtual fluorescent labeling. The difficulty of learning a domain mapping function for c^2GAN increases significantly as the number of channels grows because high-dimensional problems have larger searching space. In this part, we investigate virtual fluorescent labeling, a more challenging task for mapping a multi-channel image to another multi-channel image. Mathematically, this problem is also a regression model but involving learning a regression function between

bioRxiv preprint doi: https://doi.org/10.1101/848077; this version posted November 21, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

multivariate vectors. For the two reciprocal GANs inside c^2GAN, the backward GAN will become hard to train if the forward GAN is easy to train, and vice versa. Different from previously mentioned image coloration that establishes an implicit mapping between single-channel images and three-channel images, the task of virtual fluorescent labeling attempts to separate different structures in a z-stack composed of 13 images focused on different planes into three spectral channels.

We made the training set and the test set in a similar way as that in image coloration. Each training image collection contains 25000 instances without any aligned pairs inside. The input z-stack, network prediction, and corresponding ground truth are shown in Figs. 4a-4c, respectively. Under this set of network parameters, Most of the structures in this region can be recognized and labeled with correct fluorescent labels except the green channel is a little saturated. We also compared our results with the images generated by conventional supervised CNN[20] (see Fig. 4d), which have a more realistic intensity distribution across channels and better global effects. In this application, the supervised scheme has relatively better effects because pixel-level aligned training pairs make the model easy to train and the pixel-wise loss function can directly guide CNN to a better solution. But c^2GAN is still competitive due to its unsupervised nature. For the evaluation of each channel, some zoomed-in regions of interest (ROI) are shown in Fig. 4e-4g. We calculated the SSIM for quantitative evaluation labeling errors of each channel (Supplementary Fig. 7). Among the three channels, the blue channel has the best labeling accuracy with SSIM=0.84 and followed by the green channel with SSIM=0.78, which means that nuclei and dendrites can be resolved reliably using c^2GAN. The red channel has relatively large deviations with SSIM=0.62.

bioRxiv preprint doi: https://doi.org/10.1101/848077; this version posted November 21, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Discussion

In conventional supervised frameworks of deep learning, training a deep neural network with high generalization ability usually needs thousands of patched image pairs. To quantify the data dependency of c^2GAN, we trained c^2GAN to reconstruct super-resolution structures of microtubules with 18 unpaired training sets of different sizes. The number of patched images in these training sets ranges from 100 to 12000 and the sample points are relatively dense at the beginning to track rapid fluctuates of the network performance. The averaged normalized root-mean-square error (NRMSE, lower is better) and SSIM (higher is better) of each independent trial on the same 800 prepared test images were calculated. We also selected some representative points for a concrete interpretation of the performance (Supplementary Fig. 8). Similar to most machine learning methods, c^2GAN's performance improves as the amount of training data increases and tends to be stable when the training set size exceeds a certain level (training set size=7200, with NRMSE=0.019 and SSIM=0.84). Compared with the supervised CARE net[15] (training set size=5000, with NRMSE=0.027 and SSIM=0.85), there is no significant increase in the amount of data required for comparable performance. With the assistance of data augmentation techniques, data dependency of c^2GAN can be further weakened.

In a broader sense, pixel-wise classification can also be incorporated into the concept of pixel-wise regression. The target of pixel-wise classification is to generate two-category or multi-category labels for separating the contents from the backgrounds. In the application of typical classification tasks like segmentation, c^2GAN can solve the

bioRxiv preprint doi: https://doi.org/10.1101/848077; this version posted November 21, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

problem easily (Supplementary Notes and Supplementary Fig. 9). This fact indicates that the proposed c^2GAN is applicable in general pixel-wise regression tasks.

To summarize, we have demonstrated c^2GAN for unaligned image domain transformation in an unsupervised learning manner. With the help of cycle consistency loss, a mapping function between two image domains can be established. By enforcing an additional saliency constraint, c^2GAN can complete nonlinear pixel-wise regression problems without mapping bias, which sometimes exists in tasks involving optical microscopy. We verified the proposed framework on several image transformation tasks such as fluorescence image restoration, whole-slide histological coloration, and virtual fluorescent labeling. We made quantitative assessments to the results of c^2GAN with the reference of corresponding ground truth, as well as the results of state-of-the-art conventional CNNs trained by well-aligned image pairs. The results indicate that c^2GAN has competitive performance across different imaging modalities and various data configurations. Significantly, the non-dependence on any pre-aligned training pairs makes c^2GAN stand out. Without the laborious acquisition, annotation, and pixel-level registration, we hope that our framework can accelerate the shift in network training paradigm from a conventional supervised process to an unsupervised way, which will guide more applications of deep learning in optical microscopy.

bioRxiv preprint doi: https://doi.org/10.1101/848077; this version posted November 21, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Methods

Cycle consistency. To ensure a reliable transformation from *A* to *B*, the mapping between the source domain and the target domain should be reversible that describes a one-to-one correspondence. Hence in cycleGAN, two GANs with the same architecture are trained to complete a pair of reciprocal tasks: one GAN (the forward GAN, G) learns to convert images in *A* to *B* while the other one (the backward GAN, F) learns to do the opposite. The two GANs can finally form a closed cycle after the training process. In order to catalyze this effect, an additional term called cycle consistency loss is attached to the commonly used adversarial loss. This part quantifies the distance between the original images and the cycle-generated images. The full objective function can be formulated as follows:

$$\mathsf{L}_{cycleGAN} = \mathsf{L}_{GAN}(G) + \mathsf{L}_{GAN}(F) + \lambda \mathsf{L}_{cycle}(G,F), \tag{1}$$

where $\mathsf{L}_{GAN}(G)$ and $\mathsf{L}_{GAN}(F)$ are the adversarial losses of the forward GAN and the backward GAN, respectively. This part uses binomial cross-entropy to reflect the similarity between the generated distribution and the real distribution. $\mathsf{L}_{cycle}(G,F)$ is the cycle consistency loss that uses L1 norm to measure the reversibility of the twin-GAN system:

$$\mathsf{L}_{cycle}(G,F) = \mathbf{E}\left[\left\|F(G(a)-a\right\|_1\right] + \mathbf{E}\left[\left\|G(F(b))-b\right\|_1\right]. \tag{2}$$

This term implements restrictions on images generated by both the forward GAN and the backward GAN and ensures that the cycle-generated image should be as similar to the original image as possible. λ is a scalar to adjust the strength of cycle consistency loss.

Slackness of cycle consistency loss. Although imposing cycle consistency loss alone can achieve compelling performance in various image transformation tasks related to natural scenes, severe

bioRxiv preprint doi: https://doi.org/10.1101/848077; this version posted November 21, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

mistakes will occur when it is applied to some microscopy images, where foregrounds and backgrounds have similar profiles. In such cases, cycleGAN tends to learn a biased mapping function and produce wrong results. According to our observation, two conditions are usually required to learn a biased mapping. The first one is that the outputs of the GANs are so realistic as to fool the discriminator. This makes the twin generator-discriminator systems reach their Nash equilibrium[47], *i.e.* $L_{GAN}(G)$ and $L_{GAN}(F)$ reach optimal points. The second condition is more interesting that the two GANs are both biased. One network is biased at the output end while the other one is biased at the input end. More specifically, the forward GAN maps an original input to a biased output and then the background GAN maps the biased output back to the original input again, which still hold the cycle consistency regardless of the incorrectness of the output.

Such biased mapping can be proved to exist through a logical deduction, as shown in the top panel of Fig. 1b. Suppose an image in domain *A* and corresponding ground truth b in domain *B*, and the biased version of b is $\tilde{b}$. A concrete example is shown in the bottom panel of Fig. 1b. Without mistakes, the desired mapping functions of the two GANs are

$$G(a) = b, F(b) = a . \tag{3}$$

But when the network learns a pair of biased mappings, the forward GAN maps a to $\tilde{b}$ and the backward GAN maps $\tilde{b}$ back to a again, *i.e.*,

$$\tilde{G}(a) = \tilde{b}, \tilde{F}(\tilde{b}) = a . \tag{4}$$

The network would fail to converge to the correct mapping function because previously mentioned cycle consistency still holds well:

$$\tilde{F}(\tilde{G}(a)) = \tilde{F}(\tilde{b}) = a . \tag{5}$$

bioRxiv preprint doi: https://doi.org/10.1101/848077; this version posted November 21, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

This condition indicates that the desired mapping functions (Eq. (3)) and the biased mapping functions (Eq. (4)) are all in the optimal solution space of the loss function formulated in Eq. (1). The cycle consistency constraint alone is not tight enough. There are still some possibilities that the twin GANs will converge to some unwanted solutions.

Saliency constraint. We imposed an additional constraint on the objective function to exclude biased solutions and ensure authentic transformation. This constraint is based on the observation that, unlike most natural scenes, the backgrounds of optical microscopy images have simple and consistent forms. For instance, the background of fluorescent images is black while that of bright-field images is white. Using a simple threshold segmentation method, the rough regions of the salient objects can be extracted. Therefore, the saliency constraint is designed to be the consistency of the content masks extracted by threshold segmentation:

$$\mathsf{L}_{cons}(G,F)=\mathbf{E}\left[\left\|\mathsf{T}_{\alpha}(a)-\mathsf{T}_{\beta}(G(a))\right\|_1\right]+\mathbf{E}\left[\left\|\mathsf{T}_{\beta}(b)-\mathsf{T}_{\alpha}(F(b))\right\|_1\right], \tag{6}$$

where $\mathbf{E}$ represents element-wise averaging. T_{α} and T_{β} are segmentation operators parameterized by threshold α and β . Finally, the sum of all terms makes up of the full loss function of c^{2}GAN:

$$\mathsf{L}_{cycleGAN}=\mathsf{L}_{GAN}(G)+\mathsf{L}_{GAN}(F)+\lambda\mathsf{L}_{cycle}(G,F)+\rho\mathsf{L}_{cons}(G,F)\ , \tag{7}$$

where ρ is a hand-tuned constant to enforce saliency constraint.

Network architectures. The generator and the discriminator of c^{2}GAN mainly adopted the classical architectures of cycleGAN (Supplementary Notes and Supplementary Fig. 10). Considering the diversity of micrographs, many image transformation tasks would change the number of channels of the image, thus the two GANs must have matched input and output

bioRxiv preprint doi: https://doi.org/10.1101/848077; this version posted November 21, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

channel numbers, *i.e.*, the channel number of the output images of the forward GAN should be equal to the channel number of the output images of the backward GAN, and vice versa. For the generator, the kernel size of the first convolution layer was set to be 7 to endow this layer with large receptive fields for extracting more neighborhood information[48]. All the downsampling and upsampling layers were implemented by parametric strided convolution that could be trained for optimized image scale adjustments. The discriminator adopted PatchGAN, which applies adversarial networks in a Markovian setting for learning mappings between different image domai[49]. More specifically, it judges real or fake based on small patches in the input image and the final output label is the average of all patched results. This network configuration only penalizes structures at the scale of local patches to encourage sharp high-frequency details.

Data pre- and post- processing. In terms of the data pre-processing procedure, each dataset with ground truth was divided into training part and test part in a ratio of about 5: 1. For each image pair in the training part, we randomly decided whether to pick the original image or its ground truth with a probability of 0.5 and discard the other one. Then, the selected images were collected into domain *A* and domain *B*, respectively. This operation guaranteed that there was no aligned image pair in these two image collections (Supplementary Notes and Supplementary Fig. 11). In the test phase, all original images in the test part were fed into the pre-trained model and corresponding ground truth were reserved for comparison. Large-sized images were split into small patches with pre-defined size and overlaps were reserved for image stitching. The procedure of data processing varied in different tasks and would be further detailed in Results section if necessary.

For post-processing of the transformed images, we first allocated storage with the same size as the corresponding raw whole images and then the network output image patches were simply

bioRxiv preprint doi: https://doi.org/10.1101/848077; this version posted November 21, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

placed back to their original positions. The edge of the images beyond integral multiples of the pre-defined patch size was discarded. 3D reconstruction was performed with the built-in *3D viewer* plugin of *ImageJ*. Reconstructed data volumes were adjusted to the same perspective for the convenience of observation and comparison.

Control experiments using supervised CNNs. We benchmark our unsupervised framework against the state-of-the-art supervised CNNs to demonstrate its competitive performance. We used the CARE net[15] for the task of fluorescence image restoration. The denoising of Planaria adopted the released network architecture and parameters, and the *CSBDeep* plugin of *ImageJ* was chosen to transform the test images conveniently. For the super-resolution reconstruction of microtubules and secretory granules, we retrained the CARE net with the training set of the same size as that used in the training of c^2GAN. Another validation set composed of was prepared to monitor the whole training process and guaranteed that no overfitting occurred (Supplementary Notes and Supplementary Fig. 12). In the experiment of virtual fluorescent labeling, the elaborated-designed multi-scale CNN trained with pixel-level aligned image pairs[20] was loaded and the test images were fed to and flowed through the network. The released model, as well as the network parameters, are reliable enough to produce optimal results that the supervised network can ever achieve.

Evaluations. The evaluation strategy was on the basis of not only revealing the differences visually but also providing quantitative analyses of the transformation deviations. The main quantitative indexes that we used are PSNR, SSIM, and NRMSE, wherein PSNR and SSIM are suitable to measure high-level visual errors while NRMSE is more appropriate to quantify absolute errors related to pixel values rather than high-level structures. Especially, in the task of denoising, we calculated the distribution histograms of pixel values using *ImageJ* to indicate the

bioRxiv preprint doi: https://doi.org/10.1101/848077; this version posted November 21, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

high-fidelity performance of c^2GAN. For better visualization of the transformation deviations, we also assigned network output images and corresponding ground truth to different color channels (*i.e.*, network output images to the magenta channel, and ground truth to the green channel). Considering that magenta and green are complementary colors, if the structures before and after transformation are in the same position and with the same pixel values, these structures in the merged image will be shown in white. Otherwise, these pixels would be displayed as the color with larger pixel values (Supplementary Notes). This strategy was used to independently visualize the transformation error across different channels in whole-slide histological coloration and virtual fluorescent labeling.

Data availability

The training and test data for fluorescence image restoration are available at https://publications.mpi-cbg.de/publications-sites/7207. All relevant data used for whole-slide histological coloration are available through the Genomic Data Commons portal (https://portal.gdc.cancer.gov/). The data used during virtual fluorescent labeling can be found at https://github.com/google/in-silico-labeling/blob/master/data.md.

Code availability

Our Tensorflow implementation of c^2GAN framework is available at https://github.com/Xinyang-Li/c2GAN. The codes for data preprocessing and post-processing are also uploaded to this repository.

Acknowledgements

bioRxiv preprint doi: https://doi.org/10.1101/848077; this version posted November 21, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

We would like to acknowledge Weigert et al. for opening their source code and data on image restoration to the community. We thank the Rubin Lab at Harvard, the Finkbeiner Lab at Gladstone, and Google Accelerated Science for releasing their datasets on virtual cell staining. The pathological slides were downloaded from the Genomic Data Commons portal (https://portal.gdc.cancer.gov/) and are based on data generated by The Cancer Genome Atlas (TCGA) Program (http://cancergenome.nih.gov/). These data were publicly available without any restrictions. We thank GDC operations team for collecting, classifying and maintaining this huge amount of data, and developing user-friendly applications for data exploration.

Author Contributions

Q. D. and X. L. conceived and designed the research project. X. L., G. Z., and H. Q. designed detailed implementations and experimental schemes. X. L. and G. Z. implemented and tested the codes and performed pre- and post-processing of the data. J. W., H. X., X. L. and H. W. gave critical discussions on the results and evaluation methods. All authors participated in the writing of the paper.

Competing interests

The authors declare no competing interests.

bioRxiv preprint doi: https://doi.org/10.1101/848077; this version posted November 21, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

References

1 LeCun, Y., Bengio, Y. & Hinton, G. Deep learning. *Nature* **521**, 436-444 (2015).

2 Barbastathis, G., Ozcan, A. & Situ, G. On the use of deep learning for computational imaging. *Optica* **6**, 921-943 (2019).

3 Moen, E. et al. Deep learning for cellular image analysis. *Nat. Methods*, (2019).

4 Hornik, K., Stinchcombe, M. & White, H. Multilayer Feedforward Networks Are Universal Approximators. *Neural Networks* **2**, 359-366 (1989).

5 Belthangady, C. & Royer, L. A. Applications, promises, and pitfalls of deep learning for fluorescence image reconstruction. *Nat. Methods*, (2019).

6 Rawat, W. & Wang, Z. H. Deep Convolutional Neural Networks for Image Classification: A Comprehensive Review. *Neural Computation* **29**, 2352-2449 (2017).

7 He, K., Zhang, X., Ren, S. & Sun, J. "Deep Residual Learning for Image Recognition," in *Proc. Conference on Computer Vision and Pattern Recognition workshops* 770-778 (IEEE, 2016)

8 Nitta, N. et al. Intelligent Image-Activated Cell Sorting. *Cell* **175**, 266-276 e213,(2018).

9 Gu, Y. et al. Machine Learning Based Real-Time Image-Guided Cell Sorting and Classification. *Cytometry A* **95**, 499-509 (2019).

10 Godinez, W. J., Hossain, I., Lazic, S. E., Davies, J. W. & Zhang, X. A multi-scale convolutional neural network for phenotyping high-content cellular images. *Bioinformatics* **33**, 2010-2019 (2017).

11 Xue, D. X., Zhang, R., Feng, H. & Wang, Y. L. CNN-SVM for Microvascular Morphological Type Recognition with Data Augmentation. *J. Med. Biol. Eng.* **36**, 755-764 (2016).

12 Buggenthin, F. et al. Prospective identification of hematopoietic lineage choice by deep learning. *Nat. Methods* **14**, 403-406 (2017).

bioRxiv preprint doi: https://doi.org/10.1101/848077; this version posted November 21, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

13 Coudray, N. et al. Classification and mutation prediction from non-small cell lung cancer histopathology images using deep learning. *Nat. Med* **24**, 1559-1567 (2018).

14 Campanella, G. et al. Clinical-grade computational pathology using weakly supervised deep learning on whole slide images. *Nat. Med.* **25**, 1301-1309 (2019).

15 Weigert, M. et al. Content-aware image restoration: pushing the limits of fluorescence microscopy. *Nat. Methods* **15**, 1090-1097 (2018).

16 Ronneberger, O., Fischer, P. & Brox, T. "U-net: convolutional networks for biomedical image segmentation," in *Proc. International Conference on Medical Image Computing and Computer-Assisted Intervention*, 234–241(Springer, 2015)

17 Falk, T. et al. U-Net: deep learning for cell counting, detection, and morphometry. *Nat. Methods* **16**, 67-70 (2019).

18 Ounkomol, C., Seshamani, S., Maleckar, M. M., Collman, F. & Johnson, G. R. Label-free prediction of three-dimensional fluorescence images from transmitted-light microscopy. *Nat. Methods* **15**, 917-920 (2018).

19 Rivenson, Y. et al. Deep learning microscopy. *Optica* **4**, 1437-1443, (2017).

20 Christiansen, E. M. et al. In silico labeling: predicting fuorescent labels in unlabeled images. *Cell* **173**, 792–803 e719 (2018).

21 Goodfellow, I. et al. Generative adversarial nets. In *Proc. Advances in Neural Information Processing Systems* 2672–2680 (NIPS, 2014).

22 Isola, P., Zhu, J. Y., Zhou, T. H. & Efros, A. A. Image-to-Image Translation with Conditional Adversarial Networks. in *Proc. Conference on Computer Vision and Pattern Recognition workshops*, 1125-1134 (IEEE, 2017).

23 Wang, H. D. et al. Deep learning enables cross-modality super-resolution in fluorescence microscopy. *Nat. Methods* **16**, 103-110 (2019).

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

bioRxiv preprint doi: https://doi.org/10.1101/848077; this version posted November 21, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

25 Wu, Y. et al. Bright-field holography: cross-modality deep learning enables snapshot 3D imaging with bright-field contrast using a single hologram. *Light Sci. Appl.* **8**, 25, (2019).

26 Rivenson, Y. et al. Virtual histological staining of unlabelled tissue-autofluorescence images via deep learning. *Nat. Biomed. Eng.* **3**, 466-477(2019).

27 Zhu, J.-Y., Park, T., Isola, P. & Efros, A. A. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Proc. IEEE International Conference on Computer Vision* 2223–2232 (IEEE, 2017).

28 Zhang, Y., Liu, S., Dong, C., Zhang, X. & Yuan, Y. Multiple Cycle-in-Cycle Generative Adversarial Networks for Unsupervised Image Super-Resolution. *IEEE Trans. on Image Process.* (2019).

29 Y. Choi, M. Choi, M. Kim, J.-W. Ha, S. Kim, J. Choo, StarGAN: Unified Generative Adversarial Networks for Multi-Domain Image-to-Image Translation, in *Proc. Conference on Computer Vision and Pattern Recognition workshops*, 8789-8797 (IEEE, 2018).

30 Yi, Z., Zhang, H., Tan, P. & Gong, M. DualGAN: Unsupervised Dual Learning for Image-to-Image Translation, In *Proc. IEEE International Conference on Computer Vision* 2849-2857 (IEEE, 2017).

31 Yi, X., Walia, E. & Babyn, P. Generative adversarial network in medical imaging: A review. *Med. Image. Anal.* **58**, 101552 (2019).

32 Kang, E., Koo, H. J., Yang, D. H., Seo, J. B. & Ye, J. C. Cycle-consistent adversarial denoising network for multiphase coronary CT angiography. *Med. Phys.* **46**, 550-562 (2019).

33 You, C. et al. CT Super-resolution GAN Constrained by the Identical, Residual, and Cycle Learning Ensemble (GAN-CIRCLE). *IEEE Trans. Med. Imaging*, (2019).

34 Choi, G. et al. Cycle-consistent deep learning approach to coherent noise reduction in optical diffraction tomography. *Opt. Express* **27**, 4927-4943, (2019).

bioRxiv preprint doi: https://doi.org/10.1101/848077; this version posted November 21, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

35 Ihle, S. J. et al. Unsupervised data to content transformation with histogram-matching cycle-consistent generative adversarial networks. *Nat. Mach. Intell.* 1, 461-470 (2019).

36 Yang, W. & Yuste, R. In vivo imaging of neural activity. *Nat. Methods* **14**, 349-359(2017).

37 Sharif, S. M. A. & Jung, Y. J. Deep color reconstruction for a sparse color sensor. *Opt. Express* **27**, 23661-23681 (2019).

38 R. Zhang, P. Isola, A.A. Efros." Colorful Image Colorization," in *European Conference on Computer Vision* 649-666 (IEEE, 2016)

39 Fan, J. et al. Video-rate imaging of biological dynamics at centimetre scale and micrometre resolution. *Nat. Photonics*, (2019).

40 Royer, L. A. et al. Adaptive light-sheet microscopy for long-term, high-resolution imaging in living organisms. *Nat. Biotechnol* **34**, 1267-1278 (2016).

41 Pegard, N. C. et al. Compressive light-field microscopy for 3D neural activity recording. *Optica* **3**, 517-524 (2016).

42 Weisenburger, S. et al. Volumetric Ca(2+) Imaging in the Mouse Brain Using Hybrid Multiplexed Sculpted Light Microscopy. *Cell* **177**, 1050-1066 e1014(2019).

43 Horton, N. G. et al. In vivo three-photon microscopy of subcortical structures within an intact mouse brain. *Nat. Photonics* **7** (2013).

44 Ozeki, Y. et al. High-speed molecular spectral imaging of tissue with stimulated Raman scattering. *Nat. Photonics* **6**, 844-850 (2012).

45 Abdeladim, L. et al. Multicolor multiscale brain imaging with chromatic multiphoton serial microscopy. *Nat. Commun.* **10**, 1662(2019).

46 Mahou, P. et al. Multicolor two-photon tissue imaging by wavelength mixing. *Nat. Methods* **9**, 815-818 (2012).

47 Gharesifard, B. & Cortes, J. Distributed convergence to Nash equilibria in two-network zero-sum games. *Automatica* **49**, 1683-1692 (2013).

bioRxiv preprint doi: https://doi.org/10.1101/848077; this version posted November 21, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

48 J. Johnson, A. Alahi, and L. Fei-Fei, "Perceptual losses for real-time style transfer and super-resolution,"in *European Conference on Computer Vision* 694–711 (IEEE,2016).

49 C. Li, M. Wand. Precomputed Real-Time Texture Synthesis with Markovian Generative Adversarial Networks, in *European Conference on Computer Vision* 702-716 (IEEE,2016).

bioRxiv preprint doi: https://doi.org/10.1101/848077; this version posted November 21, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

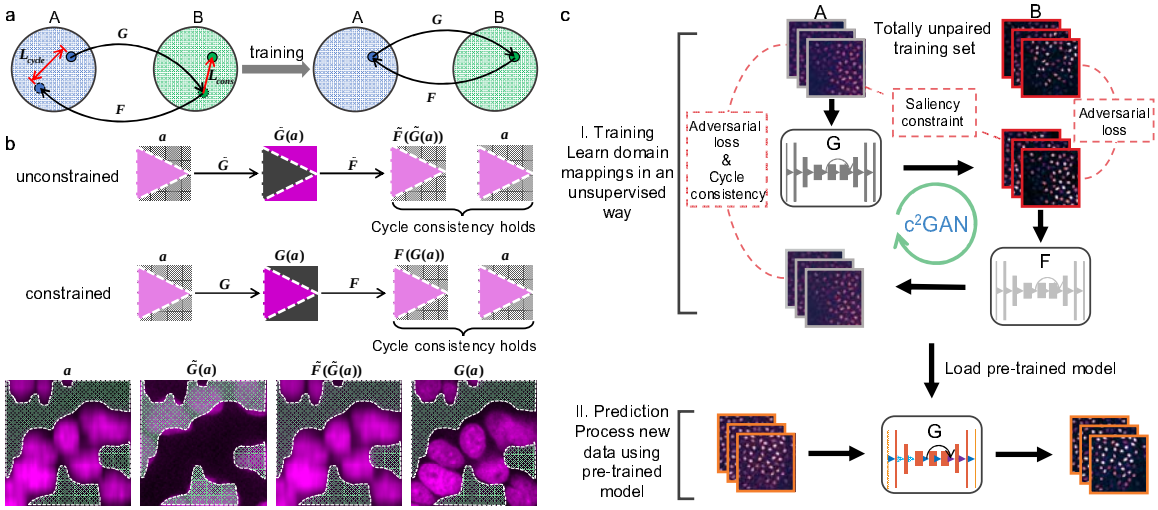

Fig. 1 | The principle and workflow of c^{2}GAN. a, Two sets *A* and *B* represent two image domains, and each domain, in fact, is a distribution in high-dimensional space and the images inside are sample points from corresponding distribution. Two GANs (G and F) with the same architecture are trained simultaneously, each for learning one direction of a pair of reciprocal mappings. A cycle-consistency constraint (L_{cycle}) along with a saliency constraint (L_{cons}) are enforced to guarantee invertibility and fidelity, respectively. After proper training, the reversible mapping between two domains can be memorized in network parameters. **b**, The cycle consistency loss alone is not strong enough to predict a satisfactory solution (G) from the input image (a) and the twin GANs is likely to converge to biased mappings ($\tilde{G}$). The saliency constraint imposes additional restrictions on cycleGAN to effectively avoid converging to biased mappings. Magenta represents the content and grey represents the background. A typical biased mapping that produces wrong results with reversed contents in absence of the saliency constraint is presented. The cycle consistency still holds well ($a \approx \tilde{F}(\tilde{G}(a))$). **c**, The

bioRxiv preprint doi: https://doi.org/10.1101/848077; this version posted November 21, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

workflow of c^2GAN includes two steps: I. Two totally unpaired image sets are prepared for training. The forward GAN (G) and the backward GAN (F) are trained to learn the mapping functions and form a closed cycle under the guidance of specially designed loss function; II. The trained network is loaded to make predictions on new data.

bioRxiv preprint doi: https://doi.org/10.1101/848077; this version posted November 21, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

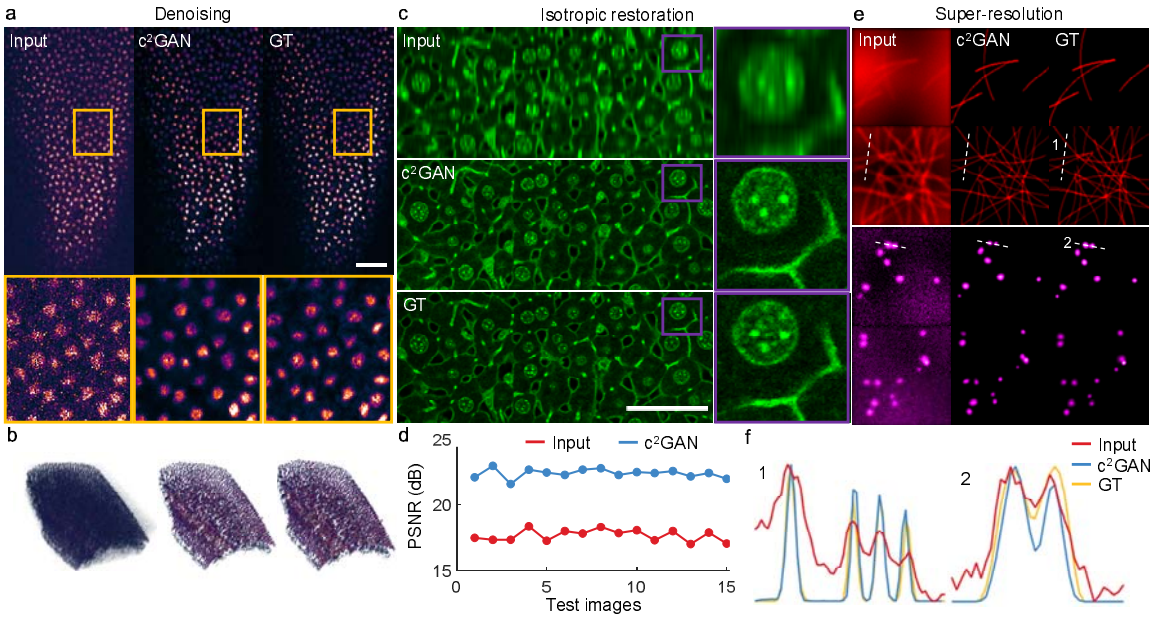

Fig. 2 | Unsupervised Image restoration with c^{2}GAN. a, c^{2}GAN learns to map low-SNR images to high-SNR images with high fidelity. The predicted images are highly consistent with ground truth (GT). Scale bar, 50 μm. **b**, A piece of surface is reconstructed to show c^{2}GAN's generalization on whole volumetric data. **c**, The degradation of axial resolution can be restored by c^{2}GAN. Scale bar, 50 μm. **d**, After the processing of c^{2}GAN, image PSNR is improved by 5dB on average. **e**, The proposed domain learning method can recover sub-diffraction structures like microtubules and secretory granules without the necessity of paired training data. **f**, Detailed intensity profiles are presented that some structures cannot be distinguished according to Rayleigh criterion can be resolved well.

bioRxiv preprint doi: https://doi.org/10.1101/848077; this version posted November 21, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

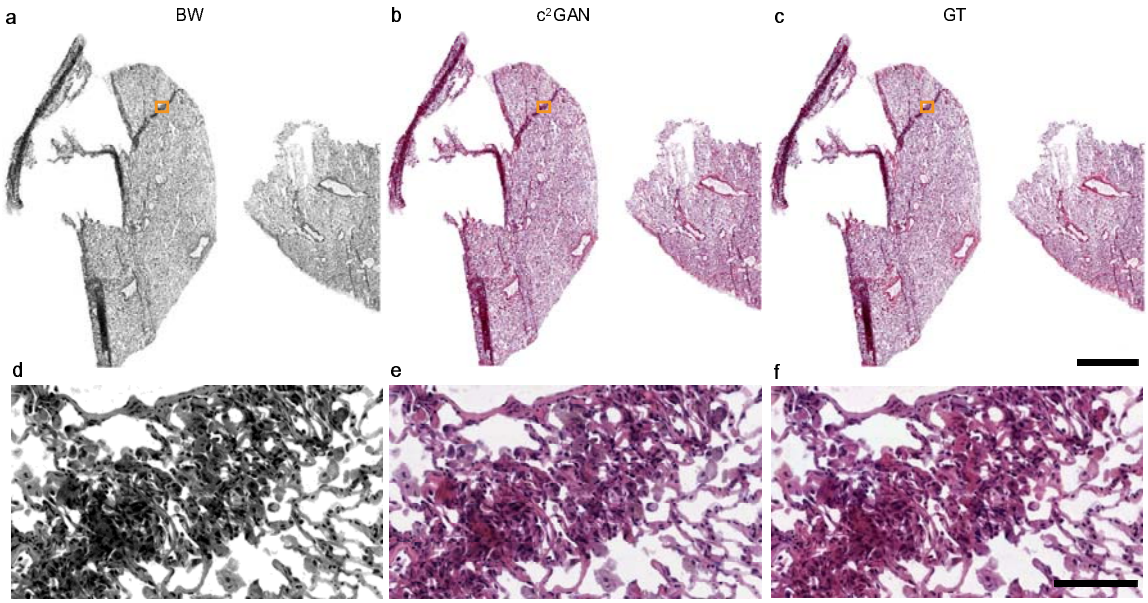

Fig. 3 | Whole-slide histological coloration using c^{2}GAN trained by totally unaligned images. a, BW version of a whole-slide pathological image. The tissue was sectioned from bronchus and lung tissue of a patient with squamous cell neoplasms. **b**, The H&E stained version of the same pathological image colored by c^{2}GAN learned from unsupervised training. **c**, The Original H&E stained image serves as the ground truth for evaluating the network prediction. Scale bar, 2 mm. **d-f,** Enlargements of the same region in **a-c**, respectively. Morphological features of the pulmonary alveoli tissue are well-reserved by c^{2}GAN. Scale bar, 100 μm.

bioRxiv preprint doi: https://doi.org/10.1101/848077; this version posted November 21, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

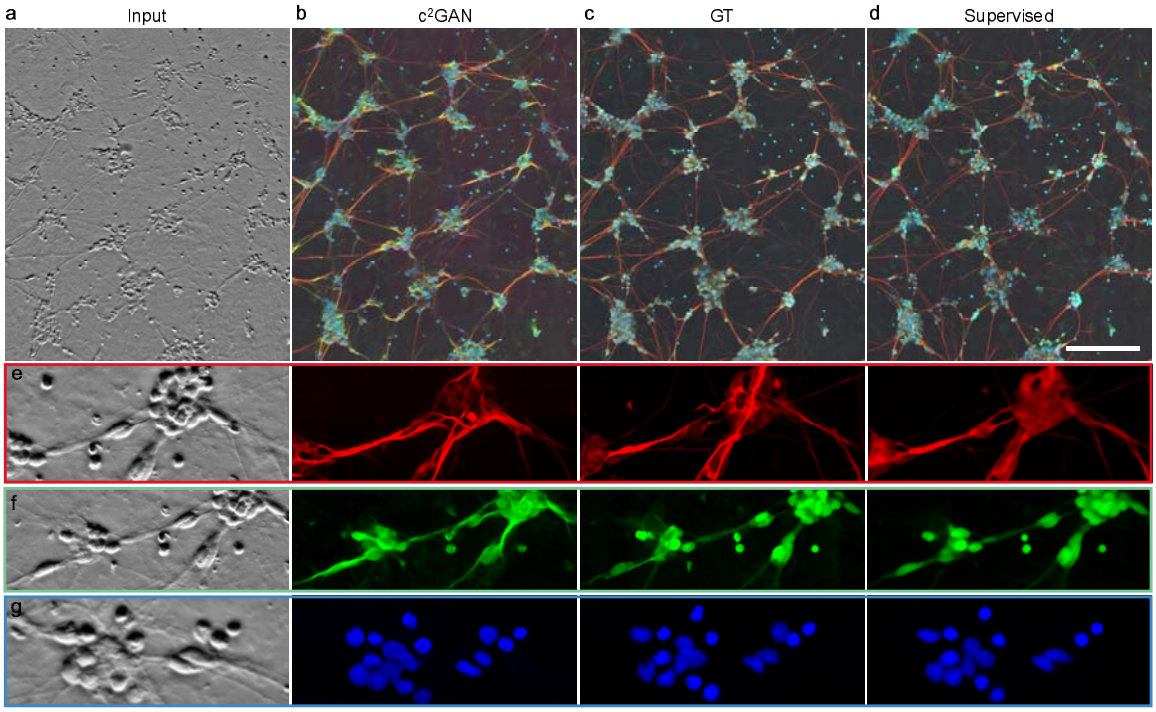

Fig. 4 | c^{2}GAN can transfer multi-depth phase-contrast images to fluorescence-labeled images in an unsupervised manner. a, The Phase contrast image of differentiated human motor neurons. The original input data cubes are z-stacks with 13 images focused on different depths, ranging from -13μm to +13 μm. The image stack is average-projected to display. **b**, The fluorescence-labeled images predicted by c^{2}GAN. **c**, The Ground truth image acquired using a spinning disc confocal microscope. Axons, dendrites and nuclei labeled by fluorescence indicators are imaged in the red, green and blue channel, respectively. **d**, The fluorescence labels predicted by the multi-scaled CNN trained by pixel-registered image pairs. Scale bar, 200 μm. **e-g**, Zoomed-in images of different ROIs split from three independent channels. The projections of axons, the profiles of dendrites, and the distribution of nuclei can be clearly revealed by c^{2}GAN.