Learning Generative Models of Tissue Organization with Supervised GANs

Ligong Han[1], **Robert F. Murphy**[2], and **Deva Ramanan**[1]

[1]Robotics Institute, Carnegie Mellon University

[2]Computational Biology Department and Department of Biological Sciences, Carnegie Mellon University

Abstract

A key step in understanding the spatial organization of cells and tissues is the ability to construct generative models that accurately reflect that organization. In this paper, we focus on building generative models of electron microscope (EM) images in which the positions of cell membranes and mitochondria have been densely annotated, and propose a two-stage procedure that produces realistic images using Generative Adversarial Networks (or GANs) in a supervised way. In the first stage, we synthesize a label "image" given a noise "image" as input, which then provides supervision for EM image synthesis in the second stage. The full model naturally generates label-image pairs. We show that accurate synthetic EM images are produced using assessment via (1) shape features and global statistics, (2) segmentation accuracies, and (3) user studies. We also demonstrate further improvements by enforcing a reconstruction loss on intermediate synthetic labels and thus unifying the two stages into one single end-to-end framework.

1. Introduction

Much research in the life sciences is now driven by large amounts of biological data acquired through high-resolution imaging [7, 19]. Such data represents an important application domain for automated machine vision analysis. Most past work has been *discriminative* in nature, focusing on trying to determine whether imaged samples differ between different patients, tissues, cell types or treatments [3, 5]. A more recent focus has been on constructing *generative* models, especially of cells or tissues [31, 26]. Such generative approaches are required in order to be able to combine spatial information on different cell types or cell organelles learned from separate images (and potentially different imaging modalities) into a single model. This is needed because of the difficulty of visualizing all components in a single image. Images can be used to perform spatially-accurate simulations of cell or tissue biochemistry [16], and synthetic images that combine many components can dramatically enhance the accuracy and usefulness of such simulations.

Reproducible Research Archive: All source code and data used in these studies is available at https://github.com/phymhan/supervised-gan.

Microscopy imaging—At the cellular scale, the dominant modes of imaging used are fluorescence microscopy (FM) and electron microscopy (EM). From the machine vision perspective, these methods differ dramatically in their resolution, noise, and the availability of labels for particular structures. FM works by tagging particular molecules or structures with fluorescence probes, adding a powerful form of sparse biological supervision to the captured images (which does not require human intervention). However, the spatial resolution of FM ranges from a limit of approximately 250 nm for traditional methods to 20–50 nm for super-resolution methods. By contrast, EM allows for significantly higher resolution (0.1–1 nm per pixel), but ability to automatically produce labels is limited and manual annotation can be very time-consuming. Analysis of EM images is also challenging because they contain lower signal-to-noise ratios than FM.

Our goal—We wish to build holistic generative models of cellular structures visible in high-resolution microscopy images. In the following, we point out several unique aspects of our approach, compared to related work from both biology and machine learning.

Data—We focus on EM images that contain enough resolution to view structures of interest. This in turns means that supervised labels (e.g., organelle segmentation masks) will be difficult to acquire. Indeed, it is quite common for standard EM benchmark datasets to contain only tens of images, illustrating the difficulty of acquiring human-annotated labels [1]. Most work has focused on segmentation of individual cells or organelles within such images [13]. In contrast, we wish to build models of *multiple* cells and their internal organelles, which is particular challenging for brain tissue due to the overlapping meshwork of neuronal cells.

Generative models—Past work on EM image analysis has focused on discriminative membrane detection [8, 15]. Here we seek a high-resolution generative model of cells and the spatial organization of their component structures. Generative models of cell organization have been a long sought-after goal [31, 27], because at some level, such models are a required component of any behavioral cell model that depends on constituent proteins within organelles.

GANs—First and foremost, we show that generative adversarial networks (GANs) [11] can be applied to build remarkably-accurate generative models of multiple cells and their structures, significantly outperforming prior models designed for FM images. To do so, we add three innovations to GANs: First, in order to synthesize large high-resolution images (similar to actual recorded EM images), we introduce *fully-convolutional* variants of GANs that exploit the spatial stationarity of cellular images. Note that such stationarity may not present in typical natural imagery (which might contain, for example, a characteristic horizon line that breaks translation invariance). Secondly, in order to synthesize natural geometric structures across a variety of scales, we add *multi-scale discriminators* to guide the generator to produce images with realistic multi-scale statistics. Thirdly, and most crucially, we make use of *supervision* to guide the generative process to produce semantic structures (such as cell organelles) with realistic spatial layouts. Much of the recent interest in generative models (at least with respect to GANs) has focused on unsupervised learning.

But in some respect, synthesis and supervision are orthogonal issues. We find that standard GANs do quite a good job of generating texture, but sometimes fail to capture global geometric structures. We demonstrate that by adding supervised structural labels into the generative process, one can synthesize considerably more accurate images than an unsupervised GAN.

Evaluation—A well-known difficulty of GANs is their evaluation. By far, the most common approach is qualitative evaluation of the generated images. Quantitative evaluation based on perplexity (the log likelihood of a validation set under the generative distribution) is notoriously difficult for GANs, since it requires approximate optimization techniques that are sensitive to regularization hyper-parameters [20]. Other work has proposed statistical classifier tests that are sensitive to the choice of classifier [17]. In our work, we use our supervised GANs to generate image-labels pairs that can be used to train discriminative classifiers, yielding quantitative improvements in prediction accuracy. Moreover, we consider the literature on generative cell models and use previously proposed metrics for evaluating generative models, including the consistency of various shape feature statistics across real vs generated images, as well as the stability of discriminative classifiers (for semantic labeling) across real vs generated data [31]. We also perform user studies to measure a user's ability to distinguish real versus generated images. Crucially, we compare to strong baselines for generative models, including established parametric shape-based models as well as non-parametric generative models that memorize the data.

2. Related Work

There is a large body of work on GANs. We review the most relevant work here.

GANs—Our network architecture is based on DCGAN [23], which introduces convolutional network connections. We make several modifications suited for processing biological data, which tends to be high-resolution and encode spatial structures at multiple scales. As originally defined, the first layer is *not* convolutional since it processes an input noise "vector". We show that by making all network connections convolutional (by converting the noise vector to a noise "image"), the entire generative model is convolutional. This in turns allows for efficient training (through learning on small convolutional crops) and high-quality image synthesis (through generation of larger noise images). We find that multi-scale modeling is crucial to synthesizing accurate spatial structures across varying scales. While past work has incorporated multi-scale cues into the generative process [9], we show that multiscale *discriminators* help further produce images with realistic multi-scale statistics.

Supervision—Most GANs work with unsupervised data, but there are variants that employ some form of auxiliary labels. Conditional GANs make use of labels to learn a GAN that synthesizes pixels conditioned on an label image [12] or image class label [21], but we use supervision to learn an end-to-end generative model that synthesizes pixels given a noise sample. Similarly, methods for semi-supervised learning with GANs [30] tend to factorize generative process into disentangled factors similar to our labels. However, such factors tend

to be global (such as an image class label), which are easier to synthesize than spatially-structured labels. From this perspective, our approach is similar to [28], who factorizes image synthesis into separate geometry and style stages. In our case, we make use of semantic labels rather than metric geometry as supervision. Finally, most related to us is [22], who uses GANs to synthesize fluorescent images using implicit supervision from cellular staining. Our work focuses on EM images, which are high resolution (and so allows for modeling of more detailed substructure), and crucially makes use of semantic supervision to help guide the generative process.

3. Supervised GANs

A standard GAN, originally proposed for unsupervised learning, can be formulated with a minimax value function $V(G,D)$:

$$\min_G \max_D V(G,D) \quad \text{where} \tag{1}$$
$$V(G,D) = \mathbb{E}_{x \sim p_x}[\log(D(x))] + \mathbb{E}_{z \sim p_z}[\log(1 - D(G(z)))]$$
$$[\textbf{UnsupervisedGAN}]$$

and x denotes image and z denotes a latent noise vector. As defined, the minimax function can be optimized with samples from the marginal data distribution p_x and thus no supervision is needed. As shown in Fig. 1, this tends to accurately generate low-level textures but sometimes fails to capture global image structures. Assume now that we have access to image labels y that specify spatial structures of interest. Can we use these labels to train a better generator? Presumably the simplest approach is to define a "classic" GAN over a joint variable $x' = (x, y)$:

$$V(G,D) = \mathbb{E}_{x,y \sim p_{xy}}[\log(D(x,y))] + \mathbb{E}_{z \sim p_z}[\log(1 - D(G(z)))] \tag{2}$$
$$[\textbf{JointGAN}]$$

Factorization—Rather than learning a generative model for the joint distribution over x, y, we can factorize it into $p(x, y) = p(y)p(x|y)$ and learn generative models for each factor. This factorization makes intuitive sense since it implicitly imposes a causal relation [14]: first geometric labels are generated with $G_y : z \mapsto y$, and then image pixels are generated conditioned on the generated labels, $G_x : y \mapsto x$. We refer to this approach as $SGAN$ (supervised GANs), as illustrated in Fig. 2-b,c:

$$V(G,D) = V_y(G_y, D_y) + V_x(G_x, D_x) \quad \text{where} \tag{3}$$
$$V_y(G_y, D_y) = \mathbb{E}_{y \sim p_y}[\log(D_y(y))] + \mathbb{E}_{z \sim p_z}[\log(1 - D_y(G_y(z)))] \quad \text{and}$$
$$V_x(G_x, D_x) = \mathbb{E}_{x,y \sim p_{xy}}[\log(D_x(x,y))] + \mathbb{E}_{y \sim p_y}[\log(1 - D_x(G_x(y), y))].$$
$$[\textbf{SGAN}]$$

In theory, one could also factorize the joint into $p(x, y) = p(x)p(y|x)$, which is equivalent to training a standard unsupervised GAN for x and a conditional model for generating labels from x. The latter can be thought of as a semantic segmentation network. We compare to such an alternative factorization in our experiments, and show that conditioning on labels first produces significantly more accurate samples of $p(x, y)$.

Optimization—Because value function $V(G,D)$ decouples, one can train $\{D_y, G_y\}$ and $\{D_x, G_x\}$ independently:

$$\min_{G} \ \max_{D} V(G, D) = \min_{G_y} \ \max_{D_y} V_y(G_y, D_y) + \min_{G_x} \ \max_{D_x} V_x(G_x, D_x) \quad (4)$$

Using arguments similar to those from [11], one can show that SGAN can recover true data distribution where the discriminator D and generators G are optimally trained:

Theorem 3.1: *The global minimum of $C(G) = \max_D V(G,D)$ is achieved if and only if $q(y) = p(y)$ and $q(x|y) = p(x|y)$, where p's are true data distributions and q's are distributions induced by G.*

Proof: Given in Supplementary A.

End-to-end learning—The above theorem demonstrates that SGANs will capture the true joint distribution over labels and data if trained optimally. However, when not optimally trained (because of optimization challenges or limited capacity in the networks), one may obtain better results through end-to-end training. Intuitively, end-to-end training optimizes $G_x(y)$ on samples of labels $\hat{y}$ produced by the initial generator $G_y(z)$, rather than ground-truth labels y. To formalize this, one can regard G_y and G_x as sub-networks of a *single* larger generator which is provided deep supervision at early layers:

$$V_x(G_x, D_x) = \mathbb{E}_{x, y \sim p_{xy}} [\ \log (D_x(x, y))] + \mathbb{E}_{z \sim p_z} [\ \log (1 - D_x(G_x(G_y(z)), G_y(z)))]\ . \quad (5)$$

However in practice, samples from an imperfect G_y makes it even harder to train G_x. Indeed, we observe that G_x produces poor results when synthetic training labels are introduced. One possible reason is that discriminator D_x will be focused on the differences between the real and predicted labels rather than correlations between labels and images. (Please refer to section 5 and Supplementary C for more analysis.) To avoid this, we force the generator to learn such correlations by also learning a *reconstructor* $F_y(x) : x \mapsto y$ that re-generates labels from images. We add a reconstruction loss $\mathcal{L}$ (similar to a "cycle GAN" [33]) that ensures that G_x will produce an image from which an accurate label can be reconstructed. We refer to this approach as a *DSGAN* (Deeply Supervised GAN):

$$\min_{G,F} \max_D V_y(G_y, D_y) + V_x(G_x, F_y, D_x) \quad \text{where} \quad (6)$$

$$V_x(G_x, F_y, D_x) = \mathbb{E}_{x \sim p_{xy}} [\log (D_x(x, y))] +$$

$$\mathbb{E}_{y \sim p_y} [\log (1 - D_x(G_x(y), y))] +$$

$$\lambda_{reg} \mathbb{E}_{x, y \sim p_{xy}} [\mathscr{L}(y, F_y(x))] +$$

$$\lambda_{cyc} \mathbb{E}_{y \sim p_y} [\mathscr{L}(y, F_y(G_x(y)))] +$$

$$\lambda_{cyc} \mathbb{E}_{z \sim p_z} [\mathscr{L}(G_y(z), F_y(G_x(G_y(z))))] .$$

[**DSGAN**]

The above training strategy is reminiscent of "teacher forcing" [29], a widely-used technique for learning recurrent networks whereby previous predictions of a network are replaced with their ground-truth values (in our case, replacing $G_y(z)$ with y). The same optimality condition as in Theorem 3.1 also holds for DSGANs.

Label editing—Another advantage of SGAN or DSGAN is label editing, because editing in label space is much easier than in image space. This allows us to easily incorporate human priors into the generating process. For example, at test time, we can perform image processing on synthetic labels such as to remove discontinuous membranes or to remove mitochondria that are concave or replace with its convex hull.

Conditional label synthesis—We can further split labels into $y = (y_1, y_2)$. This allows us to learn explicit conditionals that might be useful for simulation (e.g., synthesizing mitochondria given real cell membranes). This may be suggestive of interventions in a causal model (Causal-GAN [14]).

4. Network Architectures

In this section we outline our GAN network architectures, focusing on modifications that allow them to scale to high-resolution multi-scale biological images. Specifically, we first propose a fully-convolutional label generator which allows arbitrary output sizes; then we describe a novel multi-scale patch-discriminator to guide the generators to produce images with realistic multi-scale statistics. The fully-convolutional generator and multi-scale discriminators define a *fully-convolutional GAN* (FCGAN).

Fully-convolutional generator—Since the shape of both membrane and mitochondria are invariant to spatial location, a fully convolutional network is desirable to model the generators. Generators in previous works such as DCGAN take a noise vector as input. As a result, the size of their output images is predefined by their network architecture, thus unable to produce arbitrarily sized images at test time. We therefore propose to feed a noise "image" instead of a vector into the generator. The noise "image" is essentially a 3D tensor with the first two dimensions corresponding to the spatial positions. As illustrated in Fig. 3, to synthesize arbitrarily large labels, we only need to modify the spatial size of the input noise. Fully-convolutional generator is an instantiation of the label generator in Fig. 1-a.

Multi-scale discriminator—We initially experiment with the patch-based discriminator network as in [12], and find that the quality of synthesized labels relates to the patch size chosen for the discriminator. On the one hand, if we use a small patch size, the synthesized label has locally realistic patterns, but the global structure is wrong as it contains repetitive patterns (see Fig. 4 top-right). On the other hand, if we use a large patch-size, the output label image resembles a roughly plausible global structure, but lacks local details (see Fig. 4 bottom-right).

To ensure the generators produce both globally and locally accurate labels and images, we propose a multi-scale discriminator architecture. As illustrated in Fig. 4, the input label (or image) is first down-sampled to different scales and then fed into individual discriminators. The final discriminator output is a weighted summation of the discriminators for each scales:

$$V_y(G_y, D_y) = \mathbb{E}_{y \sim p_y} [\sum_{i \in I} \lambda_i \log (D_y^i(\pi_i(y)))] + \mathbb{E}_{z \sim p_z} [\sum_{i \in I} \lambda_i \log (1 - D_y^i(\pi_i(G_y(z))))] \quad (7)$$

Here, i is the image pyramid level index, π_i's denote down-sample transformations and λ_i's are predefined coefficients.

Conditional generator—Inspired by cascaded refinement networks (CRN) from [6], we design architecturally similar generators for both *conditional image synthesis* ($y \mapsto x$) and *conditional label synthesis* ($y_1 \mapsto y_2$). Compared to U-Net [24] which is originally adopted in pix2pix, CRN is less prone to mode collapse. Please see more discussions in Supplementary B.

5. Experiments

As illustrated in Fig. 1-a, the proposed generative process contains two parts: (1) noise $\mapsto$ label, (2) label $\mapsto$ image. Thus our generative models output both labels and images that are paired. In this paper, we also evaluate our methods on these two levels: (1) labels, and (2) images. Particularly, on the label level we locally compare the *shape features* of single cells with real ones, and globally we compute *statistics* of multiple cells. On labels, we also evaluate the *model capacity*. On the image level, we measure image qualities by *segmentation accuracy*. Also, *user studies* are conducted on both levels.

5.1. Metrics and Baselines

Shape features—Following past work [31], we evaluate the accuracy of synthetic images by (1) training a real/fake classifier, and (2) counting the portion of synthetic samples that fool the classifier. We train SVM classifiers on a set of 89 features [31] that have been demonstrated to very accurately distinguish cell patterns in FM images, and which are extracted from *label images* of single cells with mitochondria. Example statistics include 49 Zernike moment features, 8 morphological features, 5 edge features, 3 convex hull features *etc.*

Global statistics—Such shape-based features used above are typically defined for a single cell. We therefore also extracted global statistics across multiple cells, including distributions of cell size, mitochondria size and mitochondria roundness [32] *etc.*

User studies—We design an interface similar to that in [25], where generated images are presented with a prior of 50%. Intermediate labels are edited for better visual quality (samples shown in Fig. 6, cropped to 512×512).

Dataset—We used a publicly available VNC dataset [10] that contains a stack of 20 annotated sections of the Drosophila melanogaster third instar larva ventral nerve cord (VNC) captured by serial section Transmission Electron Microscopy (ssTEM). The spatial resolution is $4.6 \times 4.6 \times 50$ nm/pixel. It provides segmentation annotations for cell membranes, glia, mitochondria and synapse. Through out experiments, the first 10 sections are used for training and the remaining 10 sections are used for validation.

Parametric baseline—To construct baselines for our proposed methods, we compare to a well-established *parametric model* from [31] for synthesizing fluorescent images of single cells, which is trained by substituting nuclei with mitochondria. One noticeable disadvantage of this approach is that it lacks the capability to synthesize multiple cells. It also leverages supervision since the shape model is trained on labels.

Non-parametric baseline—We also consider a nonparametric baseline that generates samples by simple selecting a random image from the training set. We know that, by construction, such a "trivial" generative model will produce perfectly realistic samples, but importantly, those samples will not generalize beyond the training set. We demonstrate that such a generative model, while quite straightforward, presents a challenging baseline according to many evaluation measures.

5.2. Results

First, we present evaluation results on labels. For shape classifiers, example single cell labels are shown in Fig. 7 (more in Supplementary E). From Fig. 8-a, we conclude that SGAN/DSGAN and JointGAN outperform the parametric and unsupervised baselines, which is confirmed by user studies shown in Fig. 9-a. A qualitative visualization of the shape features is shown in Fig 10-a. Moreover, SGAN and DSGAN can recover the global statistics of cells as well (Table 1). Not surprisingly, shape features and global statistics extracted from the trivial non-parametric baseline model look perfect, however, the total number of different images that can be generated (which is also referred to as the support size of the generated distribution) is largely confined by the size of the dataset.

Support size—To address this limitation, we estimate the support size of the generated distribution induced by our model by computing the number of samples that need to be generated before encountering duplicates (the "Birthday Paradox" test, as proposed in past work [2]). Our model is able to produce much more diverse samples (please see Supplement D for details).

Network ablation study—As discussed, our proposed methods can produce accurate labels, which is achieved by two architectural modifications: (1) fully-convolutional generation, and (2) multi-scale discrimination. To verify their effectiveness, we conduct ablation studies, and particularly, compare FCGAN with three baselines: *Non-FC*, where G_y is not fully-convolutional; *Non-MS*, where D_y only contains a single discriminator; vanilla *DCGAN*, whose results are not shown because of poor qualities (cannot extract single cells from synthesized labels). Quantitatively, Fig. 8-b illustrates that FCGAN outperforms baselines by a large margin. A t-SNE visualization is shown in Fig. 10-b. Qualitatively, as shown in Fig. 11, Non-FC, Non-MS and DCGAN all suffer from mode collapse.

Segmentation accuracy—Following past work, we also evaluate realism of an image by the accuracy of an off-the-shelf segmentation network. We report mean IU and negative log likelihood (or NLL). Particularly, for SGAN and DSGAN, we take their image generators and use them to render images from a fixed set of pre-generated synthetic labels, which is used as "ground-truth" for evaluating segmentation accuracy. The reason is that eventually at test time, we follow the same process of rendering synthetic images from generated labels. As shown in Table 2, DSGAN has better segmentation accuracy than SGAN and Joint-GAN, which is confirmed by user studies in Fig. 9-b.

SGAN v.s. DSGAN—Perhaps it is not surprising that DSGAN performs better than SGAN, since it makes use of an additional reconstruction loss that ensures that generated images will produce segmentation labels that match (or reconstruct) those used to produce the generated images. In theory, one could add such a reconstruction loss to SGAN. However, Fig. 12 shows that SGAN+reconstructor actually has a lower mean IU (86.8%) than vanilla SGAN (87.2%). Interestingly, because SGAN explicitly factors synthesis into two distinct stages, one can evaluate the second stage module $p(x|y)$ using synthetic labels $\hat{y}$. Under such an evaluation, a reconstruction loss helps (88.3%). In fact, we found one could "game" the segmentation metric by using a pre-trained reconstructor, producing a mean IU of 93.8%. We found these generated images to be less visually-pleasing, suggesting that the generator tends to overfits to some common patterns recognized by the reconstructor.

6. Discussion

In this work, we explore methods towards supervised GAN training, where the generative process is factorized and guided by structural labels. New modifications for both generators and discriminators are also proposed to alleviate mode collapse and allow fully-convolutional generation. Finally, we demonstrate by extensive evaluation that our supervised GANs can synthesize considerably more accurate images than unsupervised baselines.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References

1. Arganda-Carreras I, Turaga SC, Berger DR, Cire an D, Giusti A, Gambardella LM, Schmidhuber J, Laptev D, Dwivedi S, Buhmann JM, et al. Crowdsourcing the creation of image segmentation algorithms for connectomics. Frontiers in neuroanatomy. 2015; 9

2. Arora S, Zhang Y. Do gans actually learn the distribution? an empirical study. 2017 arXiv preprint arXiv:1706.08224.

3. Boland MV, Murphy RF. A neural network classifier capable of recognizing the patterns of all major subcellular structures in fluorescence microscope images of hela cells. Bioinformatics. 2001; 17(12):1213–1223. [PubMed: 11751230]

4. Cao C, Liu X, Yang Y, Yu Y, Wang J, Wang Z, Huang Y, Wang L, Huang C, Xu W. , et al. Look and think twice: Capturing top-down visual attention with feedback convolutional neural networks. Proceedings of the IEEE International Conference on Computer Vision; 2015. 2956–2964.

5. Carpenter AE, Jones TR, Lamprecht MR, Clarke C, Kang IH, Friman O, Guertin DA, Chang JH, Lindquist RA, Moffat J, et al. Cellprofiler: image analysis software for identifying and quantifying cell phenotypes. Genome biology. 2006; 7(10):R100. [PubMed: 17076895]

6. Chen Q, Koltun V. Photographic image synthesis with cascaded refinement networks. 2017 arXiv preprint arXiv:1707.09405.

7. Chessel A. An overview of data science uses in bioimage informatics. Methods. 2017

8. Ciresan D, Giusti A, Gambardella LM, Schmidhuber J. Deep neural networks segment neuronal membranes in electron microscopy images. Advances in neural information processing systems. 2012:2843–2851.

9. Denton EL, Chintala S, Fergus R, et al. Deep generative image models using a laplacian pyramid of adversarial networks. Advances in neural information processing systems. 2015:1486–1494.

10. Gerhard S, Funke J, Martel J, Cardona A, Fetter R. Segmented anisotropic ssTEM dataset of neural tissue. 2013 [accessed January 26, 2018]

11. Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, Courville A, Bengio Y. Generative adversarial nets. Advances in neural information processing systems. 2014:2672–2680.

12. Isola P, Zhu J-Y, Zhou T, Efros AA. Image-to-image translation with conditional adversarial networks. 2016 arXiv preprint arXiv:1611.07004.

13. Jones C, Sayedhosseini M, Ellisman M, Tasdizen T. Neuron segmentation in electron microscopy images using partial differential equations. Biomedical Imaging (ISBI), 2013 IEEE 10th International Symposium on; IEEE; 2013. 1457–1460.

14. Kocaoglu M, Snyder C, Dimakis AG, Vishwanath S. Causalgan: Learning causal implicit generative models with adversarial training. 2017 arXiv preprint arXiv:1709.02023.

15. Lee K, Zlateski A, Ashwin V, Seung HS. Recursive training of 2d–3d convolutional networks for neuronal boundary prediction. Advances in Neural Information Processing Systems. 2015:3573–3581.

16. Loew LM, Schaff JC. The virtual cell: a software environment for computational cell biology. TRENDS in Biotechnology. 2001; 19(10):401–406. [PubMed: 11587765]

17. Lopez-Paz D, Oquab M. Revisiting classifier two-sample tests. 2016 arXiv preprint arXiv:1610.06545.

18. Maaten LvdHinton G. Visualizing data using t-sne. Journal of Machine Learning Research. Nov. 2008 9:2579–2605.

19. Meijering E, Carpenter AE, Peng H, Hamprecht FA, Olivo-Marin J-C. Imagining the future of bioimage analysis. Nature biotechnology. 2016; 34(12):1250–1255.

20. Metz L, Poole B, Pfau D, Sohl-Dickstein J. Unrolled generative adversarial networks. 2016 arXiv preprint arXiv:1611.02163.

21. Odena A, Olah C, Shlens J. Conditional image synthesis with auxiliary classifier gans. 2016 arXiv preprint arXiv:1610.09585.

22. Osokin A, Chessel A, Salas REC, Vaggi F. Gans for biological image synthesis. ICCV. 2017

23. Radford A, Metz L, Chintala S. Unsupervised representation learning with deep convolutional generative adversarial networks. 2015 arXiv preprint arXiv:1511.06434.

24. Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation. International Conference on Medical Image Computing and Computer-Assisted Intervention; Springer; 2015. 234–241.

25. Salimans T, Goodfellow I, Zaremba W, Cheung V, Radford A, Chen X. Improved techniques for training gans. Advances in Neural Information Processing Systems. 2016:2234–2242.

26. Svoboda D, Homola O, Stejskal S. Generation of 3d digital phantoms of colon tissue. International Conference Image Analysis and Recognition; Springer; 2011. 31–39.

27. Svoboda D, Kozubek M, Stejskal S. Generation of digital phantoms of cell nuclei and simulation of image formation in 3d image cytometry. Cytometry part A. 2009; 75(6):494–509.

28. Wang X, Gupta A. Generative image modeling using style and structure adversarial networks. European Conference on Computer Vision; Springer; 2016. 318–335.

29. Williams RJ, Zipser D. A learning algorithm for continually running fully recurrent neural networks. Neural computation. 1989; 1(2):270–280.

30. Zhang H, Deng Z, Liang X, Zhu J, Xing EP. Structured generative adversarial networks. Advances in Neural Information Processing Systems. 2017:3900–3910.

31. Zhao T, Murphy RF. Automated learning of generative models for subcellular location: building blocks for systems biology. Cytometry Part A. 2007; 71(12):978–990.

32. Zheng J, Hryciw R. Traditional soil particle sphericity, roundness and surface roughness by computational geometry. Géotechnique. 2015; 65(6):494–506.

33. Zhu JY, Park T, Isola P, Efros AA. Unpaired image-to-image translation using cycle-consistent adversarial networks. 2017 arXiv preprint arXiv:1703.10593.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript
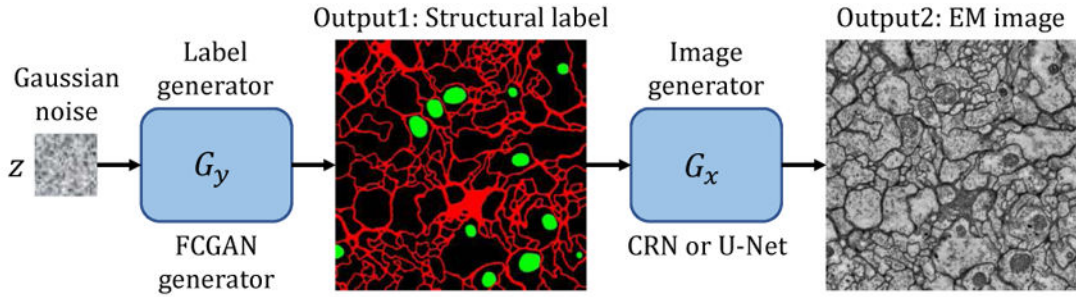
(a) Generative pipeline
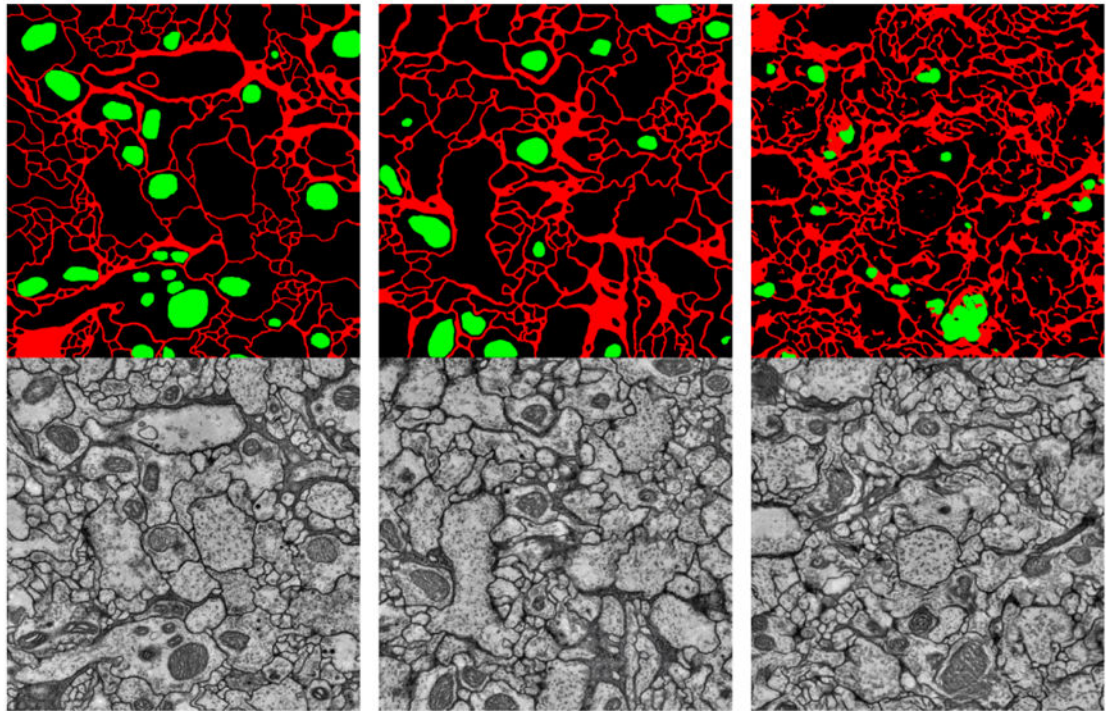
(b) Ground-truth   (c) SGAN (ours)   (d) UnsupervisedGAN

Figure 1.
(a) Generative pipeline: Given noise "image" z sampled from a Gaussian distribution, our label generator G_y generates a label image, which is then translated into an EM image by G_x. (b) Ground-truth label-image pair. (c) Label and image pair generated by our supervised GANs (SGAN), that is capable of generating continuous membranes (red lines) and correctly positioned mitochondria (green blobs). (d) Image synthesized by unsupervised GANs, in which the label is generated by a pre-trained semantic segmentation network. Unsupervised GAN is able to produce pixel-level details locally but fails to capture structures globally.
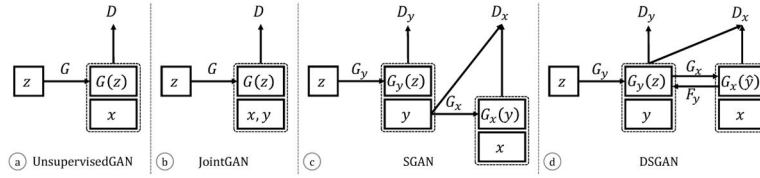
Figure 2.

We compare different GAN architectures for injecting supervision (provided with labels y) into a generative model of x. (a) A standard unsupervised GAN for generating x. (b) A GAN defined over a joint variable $x' = (x, y)$. (c) SGAN, which is a supervised GAN that is composed of an initial GAN $\{G_y, D_y\}$ that generates labels y followed by a conditional GAN $\{G_x, D_x\}$ that generates images x from y. (d) A Deeply supervised GAN (DSGAN), equivalent to a single GAN that is provided deep supervision for generating labels at an intermediate stage. Performance is further improved by adding a reconstructor F_y that ensures that generated images can be used to predict labels with low reconstruction error (Eq. 6).
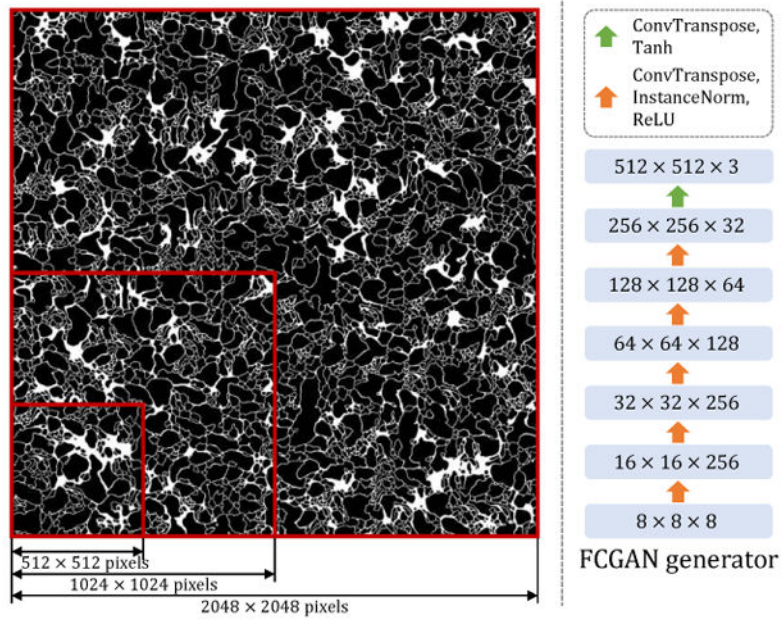
Figure 3.
A fully-convolutional generator G_y (FCGAN). Left: By changing the size of the input noise "image", our FCGAN generator can synthesize arbitrarily large labels. Right: Architecture of the fully-convolutional generator.
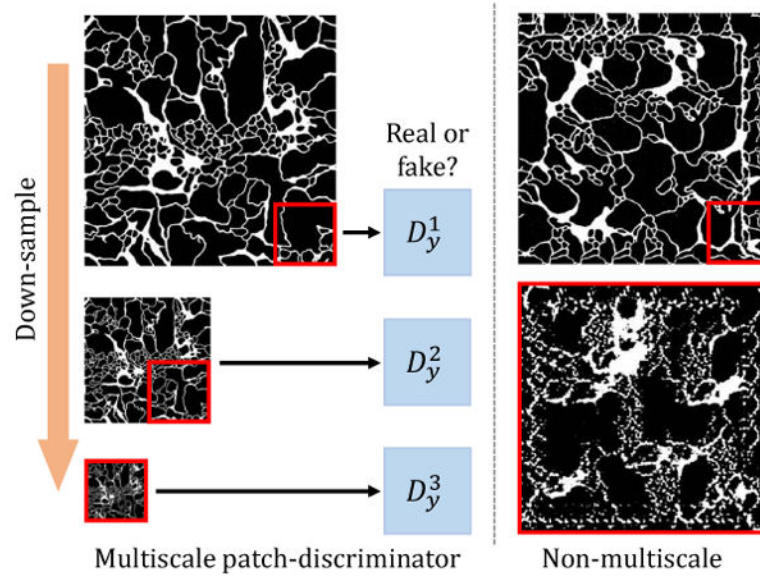
Figure 4.
Multi-scale discriminators for D_y and D_x. Left: We construct an image pyramid from the generated label (or image), and feed patches from this pyramid to multiple patch-based discriminators. Right: Single-scale discriminators with small receptive fields (top) tend to produce accurate local structure, but inaccurate repetitive global structure. Similarly, single-scale discriminators with large receptive fields produce accurate global structure, but fail to generate accurate local textures.
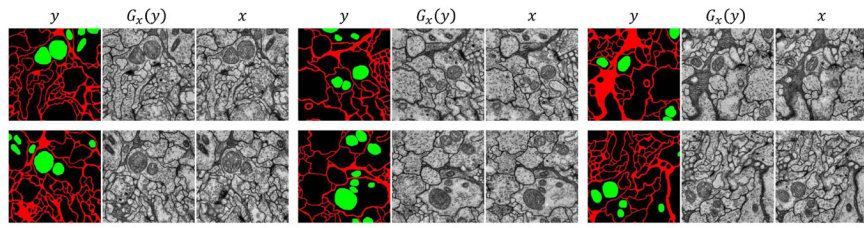
Figure 5.
Conditional image synthesis. Given true label y, we sample image $G_x(y)$, compared to real image x.

(a) Synthesized label-image pairs by SGAN.

(b) Synthesized label-image pairs by JointGAN.   (c) Synthesized label-image pairs by UnsupervisedGAN.
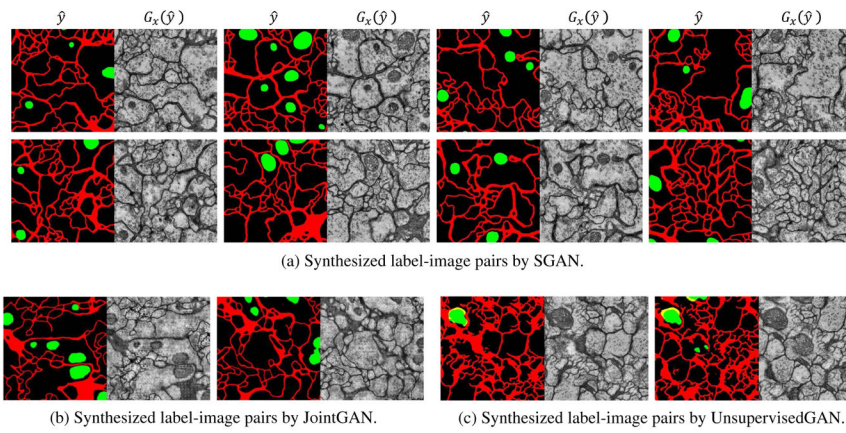
Figure 6.
Samples of our full model. (a)We first sample synthetic label $\hat{y} = G_y(z)$ from noise z, then generate image $G_x(\hat{y})$. We perform label editing (remove discontinuous membranes and concave mitochondria) on synthetic labels. (b) Label-image pairs directly synthesized by FCGAN. (c) Images are first generated by FCGAN then labels are inferred by an off-the-shelf segmentation network. Some pixels are labeled in yellow because because we use two separate segmentation networks for membranes and mitochondria.

(a) Non-Parametric   (b) Parametric
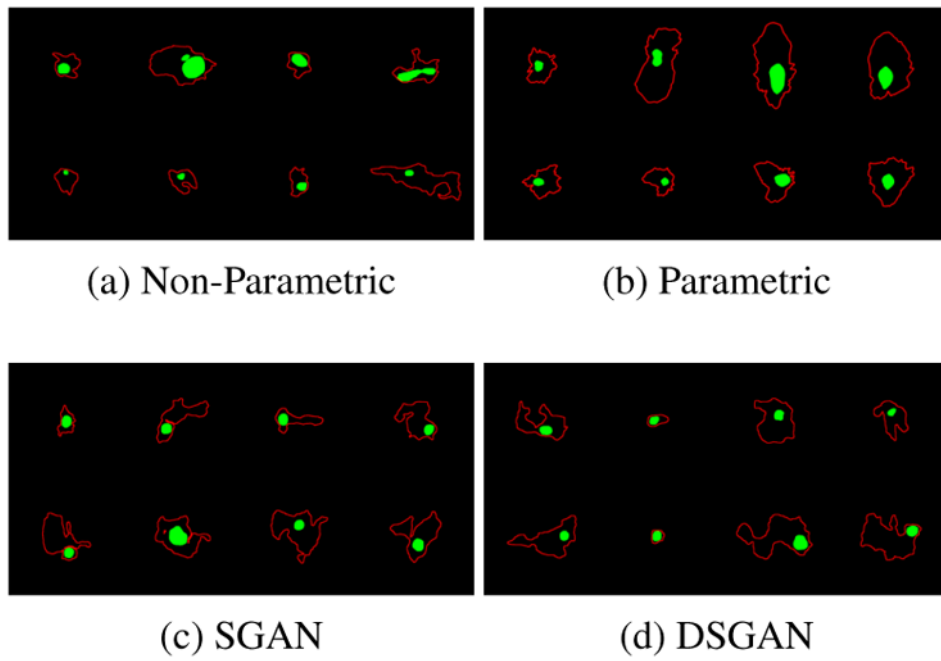
(c) SGAN   (d) DSGAN

Figure 7.
Samples of single cell labels, from on which biological markers are extracted. More samples for other baselines are given in Supplementary E. Quantitative results can be found in Fig. 8(a) (b).
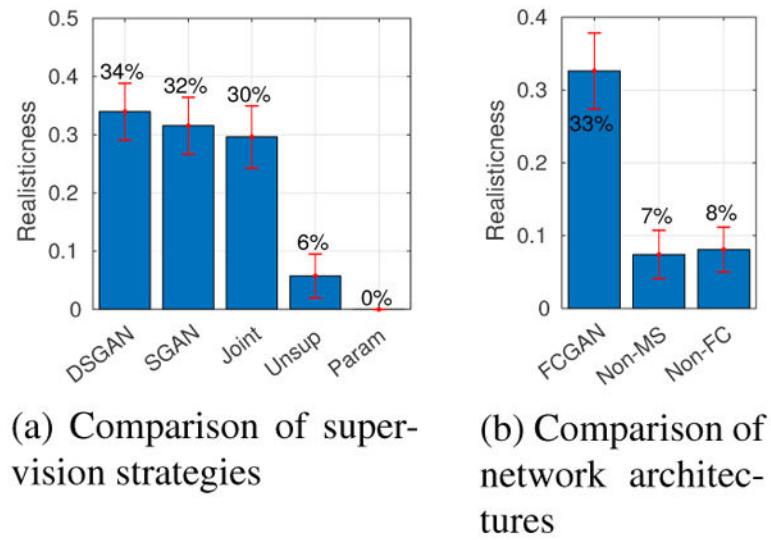
(a) Comparison of supervision strategies

(b) Comparison of network architectures

Figure 8.
Evaluating realism with image classifiers. SVM classifiers are trained to distinguish real and synthetic single-cell labels based on shape features as described in section 5.1. Bar plot shows percentage of synthetic labels being classified as real ones, higher is better.
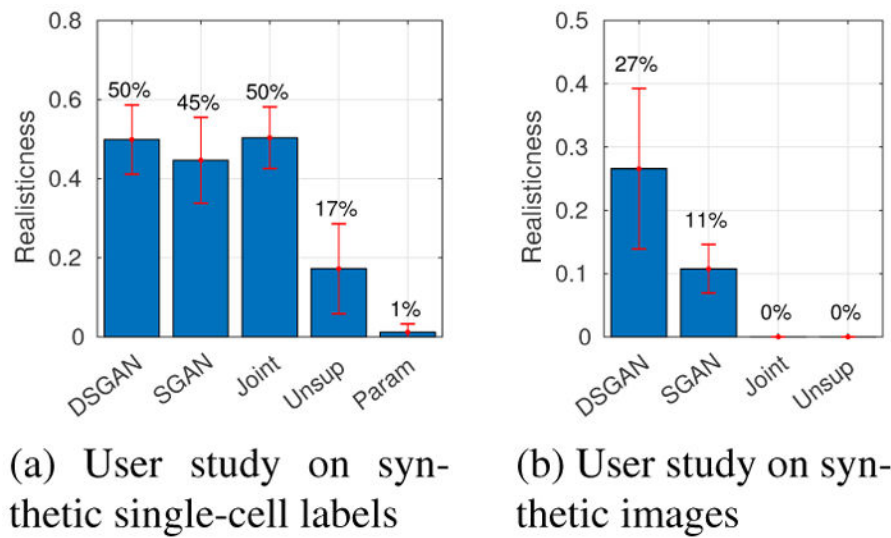
(a) User study on synthetic single-cell labels

(b) User study on synthetic images

Figure 9.
Evaluating realism with user studies. Bar plot shows percentage of synthetic images being classified by users as real ones, higher is better. (a) Users are shown mixtures of real and synthetic labels of single cells. (b) Users are shown mixtures of real and synthetic full images.
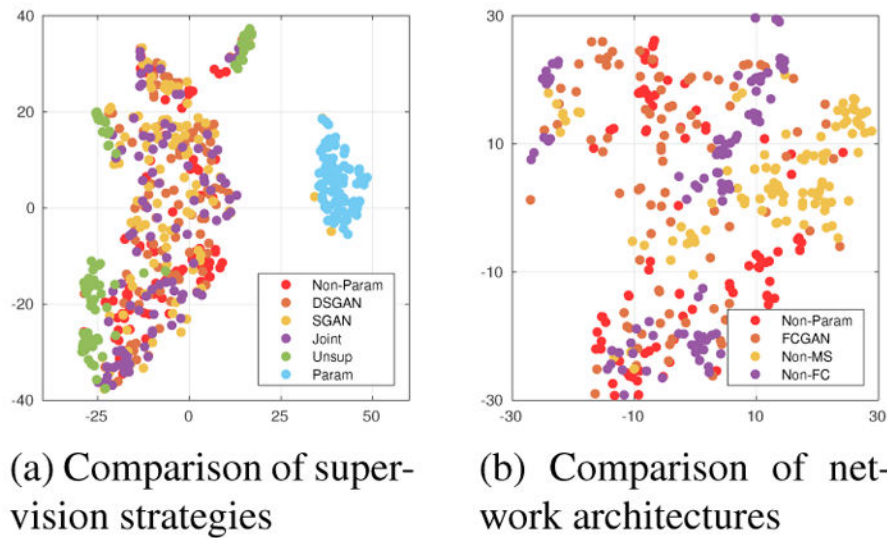
(a) Comparison of super-vision strategies

(b) Comparison of net-work architectures

Figure 10.
2-D t-SNE [18] visualization of the 89-dimensional shape features. (a) Features of DSGAN, SGAN and JointGAN well overlap with real ones (Non-Parametric baseline), while features of Unsupervised-GAN or parametric baseline are easily separable. (b) Non-FC and Non-MS only covers parts of the real (projected) feature distribution.
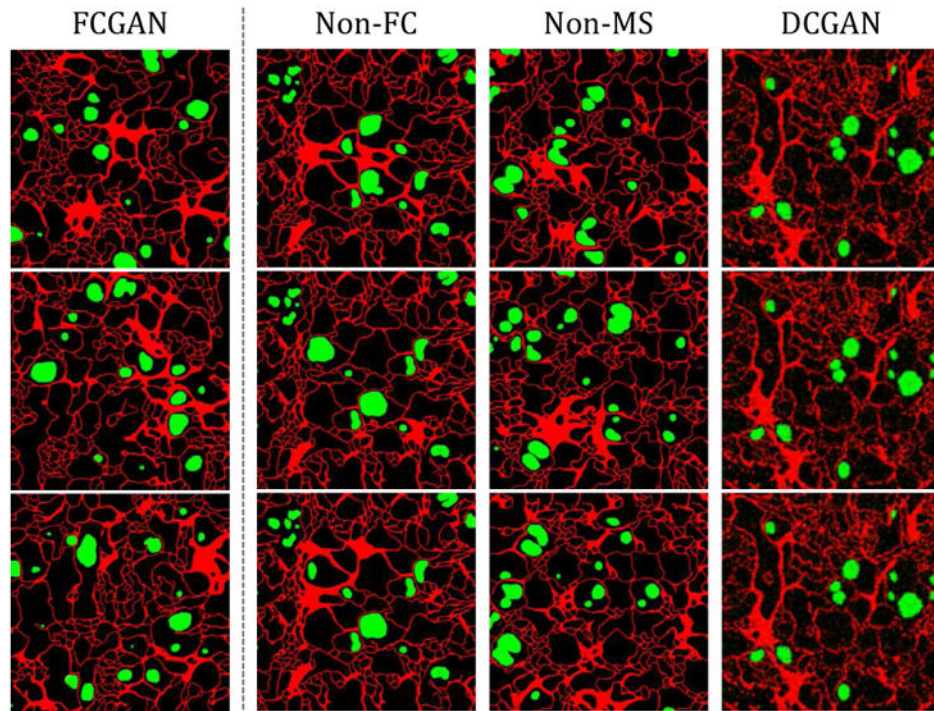
Figure 11.
Label synthesis, raw output without label editing. Non-FC, Non-MS and DCGAN all suffer from mode collapse: Non-FC, patterns at four sides are the same across samples, inner patterns are also repetitive; Non-MS, repetitive patterns show at different locations; DCGAN, samples are blurry and almost identical.
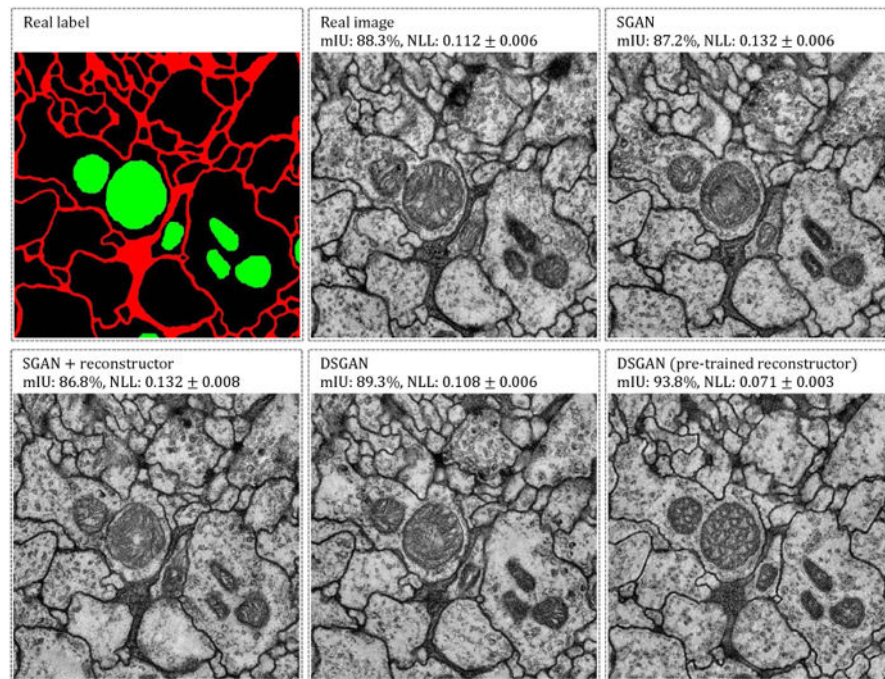
Figure 12.
Synthetic image samples and segmentation accuracies of different training approaches. We take G_x's and evaluate segmentation accuracies on a same set of generated labels. G_x of DSGAN yields higher segmentation accuracy but does not show obvious advantage visually. DSGAN with a pre-trained reconstructor achieves the highest score but not in terms of visual inspection.

Table 1

Global statistics of multiple cells. We report the numbers of average cell size, average mitochondria size, average mitochondria roundness, average number of mitochondria per cell, and chi-squared distances of their distributions between ground-truth and synthetic data. Numbers suggest SGAN/DSGAN captures global statistics of cellular structures.

	avg cell size (μm)	avg mito size (μm)	avg mito roundness	avg #of mito per cell	χ^2 cell size	χ^2 mito size	χ^2 mito roundness
Training	0.286	0.291	0.843	0.067	0	0	0
Non-Param	**0.283**	0.270	0.797	0.074	0.054	**0.094**	0.103
DSGAN	0.310	0.272	0.819	**0.068**	**0.050**	0.130	0.109
SGAN	0.311	**0.288**	**0.839**	0.073	0.057	0.112	0.122
Joint	0.272	0.275	0.779	0.060	0.058	0.129	**0.095**
Unsup	0.208	0.231	0.646	0.017	0.158	0.424	0.270

Table 2

Segmentation accuracies for SGAN/DSGAN and baselines. The mean IU and NLL of SGAN/DSGAN both match those of real cell images. Non-FC and Non-MS have high segmentation accuracy due to mode collapse. Unsupervised-GAN is not shown because it does not provide "ground-truth" label automatically

Dataset	mean IU	NLL
Non-Param	88.3%	0.112 ± 0.006
DSGAN	89.3%	0.108 ± 0.006
SGAN	87.2%	0.132 ± 0.006
Joint	81.8%	0.177 ± 0.013