

# UC Santa Cruz

## UC Santa Cruz Previously Published Works

### Title

Enhanced cell segmentation with limited training datasets using cycle generative adversarial networks

### Permalink

<https://escholarship.org/uc/item/84k261kg>

### Journal

iScience, 27(5)

### ISSN

2589-0042

### Authors

Zargari, Abolfazl  
Topacio, Benjamin R  
Mashhadi, Najmeh  
[et al.](#)

### Publication Date

2024-05-01

### DOI

10.1016/j.isci.2024.109740

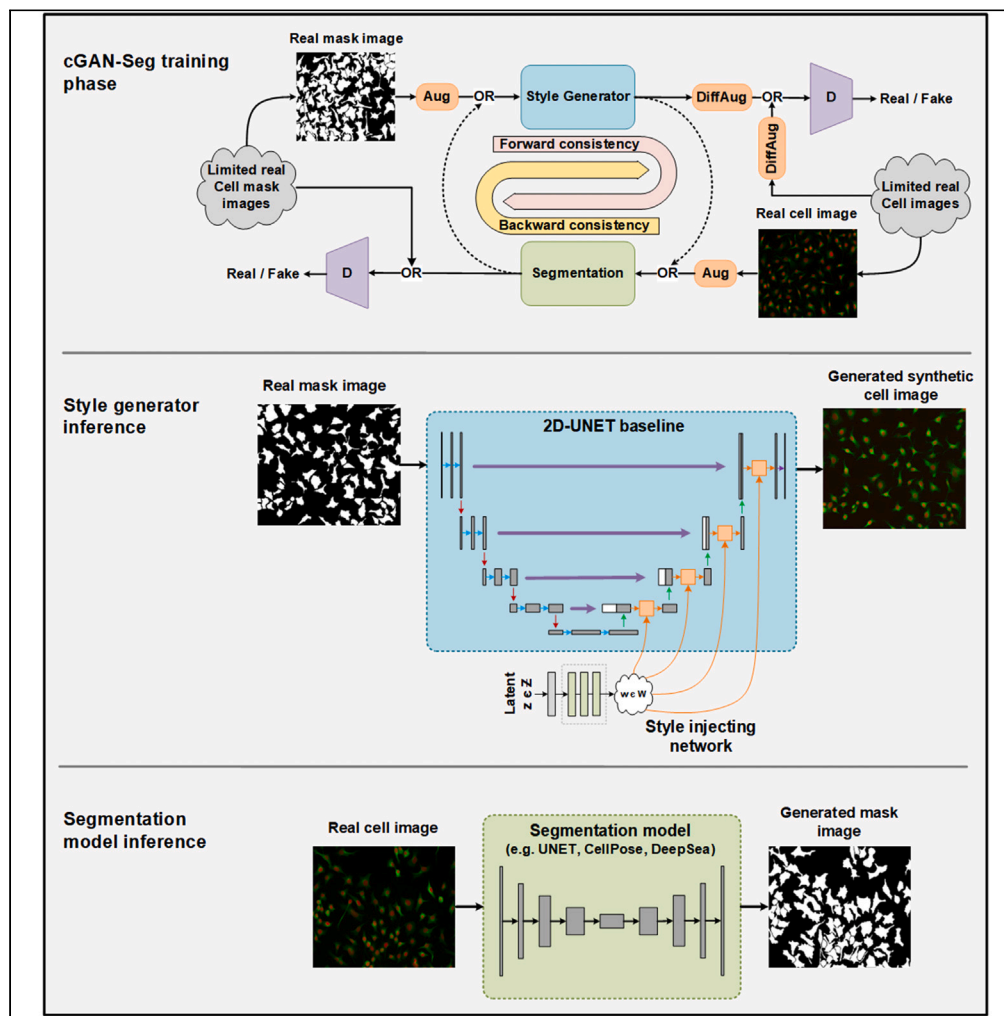
### Copyright Information

This work is made available under the terms of a Creative Commons Attribution-NonCommercial-NoDerivatives License, available at <https://creativecommons.org/licenses/by-nc-nd/4.0/>

Peer reviewed

## Article

# Enhanced cell segmentation with limited training datasets using cycle generative adversarial networks



Abolfazl Zargari,  
Benjamin R.  
Topacio, Najmeh  
Mashhadi, S. Ali  
Shariati

abzargar@ucsc.edu (A.Z.)  
alish@ucsc.edu (S.A.S.)

## Highlights

cGAN-Seg is a customized cycleGAN model designed for enhanced cell segmentation

cGAN-Seg generates realistic synthetic phase-contrast and fluorescent images

cGAN-Seg improves the accuracy of segmentation models with limited annotated data

cGAN-Seg works across various cell types and widely used segmentation models

Zargari et al., iScience 27,  
109740  
May 17, 2024 © 2024 The  
Author(s). Published by Elsevier  
Inc.  
[https://doi.org/10.1016/  
j.isci.2024.109740](https://doi.org/10.1016/j.isci.2024.109740)

## Article

## Enhanced cell segmentation with limited training datasets using cycle generative adversarial networks

Abolfazl Zargari,<sup>1,\*</sup> Benjamin R. Topacio,<sup>3,4,5</sup> Najmeh Mashhadi,<sup>2</sup> and S. Ali Shariati<sup>3,4,5,6,\*</sup>

## SUMMARY

Deep learning is transforming bioimage analysis, but its application in single-cell segmentation is limited by the lack of large, diverse annotated datasets. We addressed this by introducing a CycleGAN-based architecture, cGAN-Seg, that enhances the training of cell segmentation models with limited annotated datasets. During training, cGAN-Seg generates annotated synthetic phase-contrast or fluorescent images with morphological details and nuances closely mimicking real images. This increases the variability seen by the segmentation model, enhancing the authenticity of synthetic samples and thereby improving predictive accuracy and generalization. Experimental results show that cGAN-Seg significantly improves the performance of widely used segmentation models over conventional training techniques. Our approach has the potential to accelerate the development of foundation models for microscopy image analysis, indicating its significance in advancing bioimage analysis with efficient training methodologies.

## INTRODUCTION

Generative adversarial networks (GANs) have gained significant attention in recent years due to their remarkable success in generating realistic images and videos.<sup>1,2</sup> GANs are deep learning architectures that consist of two neural networks: a generator and a discriminator.<sup>3–5</sup> The generator network is responsible for synthesizing new data, while the discriminator network attempts to distinguish between real and synthetic samples. The two networks compete in a game-like scenario until the generator produces data that is nearly indistinguishable from real data. GANs have exhibited substantial capabilities across a wide variety of applications, including image synthesis,<sup>1</sup> video generation,<sup>6</sup> and natural language processing.<sup>7</sup>

Cell segmentation is a crucial step in microscopy, which involves the identification and delineation of individual cells within images. Cell segmentation is inherently complex due to the diversity and irregularity of morphological features of different cell types, such as shape and size, as well as the propensity for cells to cluster together, making highly accurate segmentation a challenging task. Deep learning methods, particularly convolutional neural networks (CNNs), have shown great success in improving cell segmentation accuracy.<sup>8–12</sup> The development and application of deep learning models for cell segmentation heavily depend on the availability of a large amount of annotated training data.

While recent efforts have introduced useful large datasets of microscopy images such as LiveCELL,<sup>13</sup> the manual annotation of cell images remains a laborious and time-consuming task. Although software such as Cell-ACDC and microSAM have made significant improvements in segmentation, annotation speed, and efficiency by assisting experts in the manual labeling process,<sup>9,14</sup> outlining individual cells in microscopic images to create “ground truth” masks still requires considerable effort by cell biology experts. This process becomes practically infeasible when dealing with large volumes of data or when timely results are needed. Despite the progress in dataset development, the scarcity of annotated data still poses challenges to training robust models, thus slowing down progress in building cell segmentation models that work across diverse imaging scenarios and modalities.

In this study, we propose a solution to this challenge by employing a CycleGAN-based segmentation model, termed cGAN-Seg, that is designed to train cell segmentation models with limited annotated data. We used cGAN-Seg to generate realistic and diverse microscopy images of cells in different modalities, thereby enriching the training data for the segmentation model. We systematically compared the cGAN-Seg training approach with that of the conventional training approach in terms of the performance of our segmentation model. Our results showed that the cGAN-Seg approach, with limited annotated cell image datasets, increased the model’s ability to generalize and, thus, enhanced the performance of cell segmentation tasks.

<sup>1</sup>Department of Electrical and Computer Engineering, University of California, Santa Cruz, Santa Cruz, CA, USA<sup>2</sup>Department of Computer Science and Engineering, University of California, Santa Cruz, Santa Cruz, CA, USA<sup>3</sup>Department of Biomolecular Engineering, University of California, Santa Cruz, Santa Cruz, CA, USA<sup>4</sup>Institute for The Biology of Stem Cells, University of California, Santa Cruz, Santa Cruz, CA, USA<sup>5</sup>Genomics Institute, University of California, Santa Cruz, Santa Cruz, CA, USA<sup>6</sup>Lead contact\*Correspondence: [abzargar@ucsc.edu](mailto:abzargar@ucsc.edu) (A.Z.), [alish@ucsc.edu](mailto:alish@ucsc.edu) (S.A.S.)<https://doi.org/10.1016/j.isci.2024.109740>

## RESULTS

### Designing an enhanced cycle-generative adversarial network architecture

An overview of our cGAN-Seg design is illustrated in Figure 1A. The cGAN-Seg architecture uses a customized CycleGAN approach<sup>15</sup> for cell image segmentation, offering enhanced data diversity. The cGAN-Seg has several new key features: (i) Application of a proposed style generation path within a 2D-UNET-based image generator,<sup>16,17</sup> our approach captures complex image variations from varied cell shapes to intricate structures and textures and also enhances the creation of diverse synthetic microscopy cell images (Figure S2). This prepares the model to tackle a wide range of image conditions, boosting its versatility and accuracy, (ii) Two adversarial PatchGAN discriminators enhanced with a linear attention layer,<sup>18,19</sup> our model focuses on the most salient features, producing synthetic images that closely mirror real samples (Figure S3). This approach ensures stable and efficient training dynamics, making the synthetic images more authentic, (iii) Implementation of a differentiable image augmentation technique<sup>20</sup> in our design applies identical differentiable augmentations to both real and synthetic samples (Figures 1A and S1). This reduces the risk of the discriminator memorizing exact training samples, combats overfitting, and improves the realism of synthetic images, aligning them closely with real-world scenarios, and finally (iv) Through a balanced utilization of different loss functions, including L1 and VGG-based perceptual loss functions (for the generation model)<sup>21,22</sup> and Cross-Entropy and Dice losses (for the segmentation model), the model ensures the production of high-fidelity synthetic cell images while maintaining segmentation accuracy (Equations 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, and 17).

### Model's performance evaluation

To test our proposed cGAN-Seg approach, we designed and ran three distinct segmentation training scenarios: 1) Employing cGAN-Seg to train segmentation models on a selected small subset of a dataset, followed by evaluation on the selected test set (Figure 1A); 2) Direct and conventional training of segmentation models on the same small subset without cGAN-Seg, with subsequent testing on the same test set (Figure 1B); 3) Direct and conventional training of segmentation models on the full dataset of the same cell type and assessing its performance on the same test set (Figure 1C). To ensure the robustness and generalizability of our proposed approach across a diverse range of biological contexts, we utilized samples from four different image datasets, each featuring different modalities and cell types. These datasets include the DeepSea Dataset,<sup>12</sup> the LiveCell dataset,<sup>13</sup> the Cell Tracking Challenge dataset,<sup>23</sup> and the CellPose dataset.<sup>11</sup> For the segmentation tasks, we used the DeepSea baseline architecture,<sup>12</sup> as well as two of the most widely utilized baselines in the field: the 2D-UNET<sup>17</sup> and the CellPose models<sup>11</sup>; The 2D-UNET is known for its high performance and suitability for biomedical imaging due to its encoder-decoder architecture, the DeepSea architecture is a streamlined adaptation of 2D-UNET optimized for simplicity without compromising on efficiency; and CellPose uses a versatile neural network that is trained on diverse cell types, allowing for generalization across various cell types.

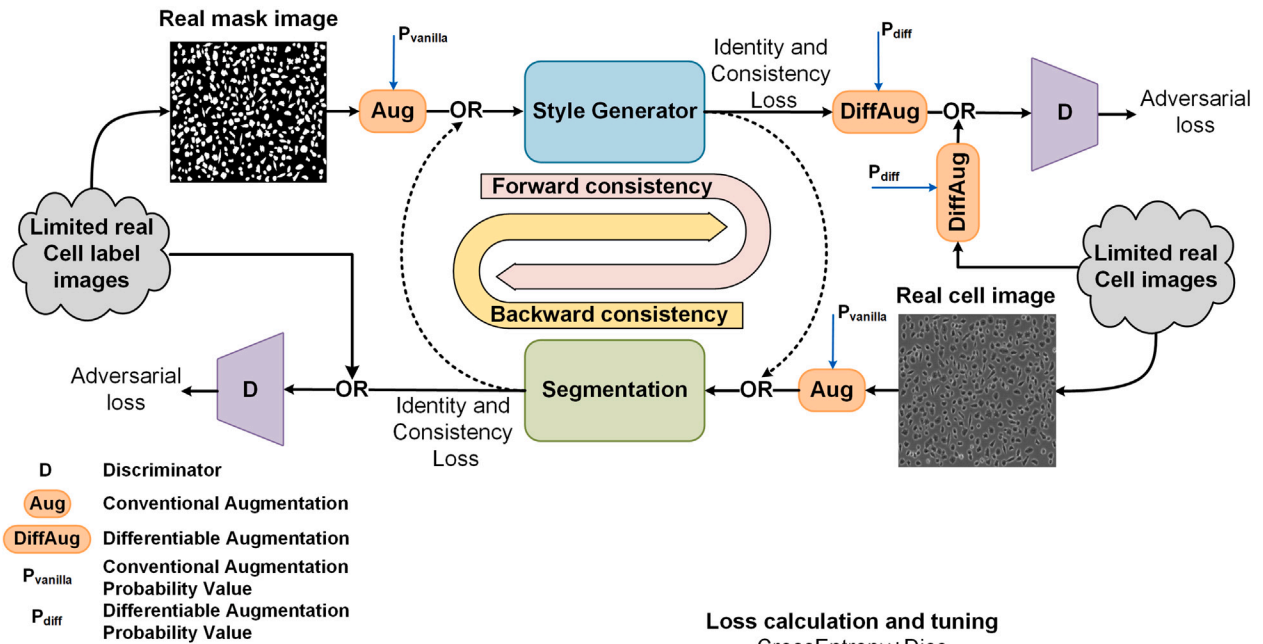
In the testing phase, our evaluation focused on the accuracy of the segmentation models by measuring the Intersection Over Union (IOU) between predicted and true cell masks. Cells detected with an IOU at or above our threshold were labeled as the correct detection. To provide a holistic view of model performance, we calculated precision, recall, and the F-score. These metrics inherently account for the impact of false positives (incorrectly predicted cells) and false negatives (missed actual cells), offering a balanced measure of our models' effectiveness across varying IOU thresholds (Equations 2, 3, and 4). In all experiments and score reports, we applied the 5-fold cross-validation technique, aiming to provide a more robust assessment of the segmentation model's performance by reducing the impact of random variations in the training and validation data splits and ensuring that the model's performance is not overly influenced by a specific subset of the data.

Figure 2 presents the average f-score of three segmentation models trained using the cGAN-Seg method on a small dataset of 200 samples. This performance is compared with that achieved by conventional training methods (Figures 1B and 1C), using both the same small dataset and a larger dataset of 1000 samples. It focuses on two distinct cell types from the DeepSea dataset for which sufficiently large, annotated image collections were available. In all test experiments, cGAN-Seg significantly improved the performance of three segmentation models of 2D-UNET, CellPose, and DeepSea almost at all IOU threshold values compared to when we trained the segmentation model with the same limited dataset. This enhancement is particularly evident in stem cell samples, which pose a greater challenge due to the higher diversity and complexity of cell images. Importantly, the segmentation scores achieved using the cGAN-Seg method with a limited dataset closely match, or even surpass, those of segmentation models trained with traditional approaches on large datasets. This outcome underscores the cGAN-Seg architecture's style generator's capability to produce a diverse enough array of samples during training. This diversity effectively compensates for the limited data available to the segmentation model, showcasing the method's efficiency in optimizing performance despite data constraints.

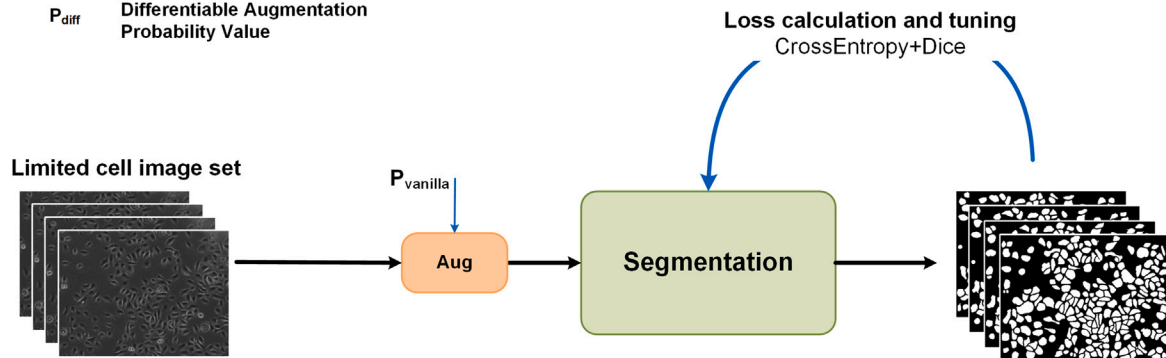
Next, we extended our analysis to two other publicly available datasets or microscopy images, LiveCell and Cell Tracking Challenge, to test the applicability of cGAN-Seg for a diverse set of cell types. Table 1 summarizes an f-score comparison between the conventional segmentation and our proposed cGAN-Seg training method for different cell types in the LiveCell and Cell Tracking Challenge datasets. In addition, we measured the average recall, precision, and dice score metrics (Tables S1–S3) for the 2D-UNET and CellPose models. The results demonstrate that the cGAN-Seg model improves the segmentation model's performance in every case. This consistent enhancement across different cell types and experimental conditions highlights the model's robustness and its adaptability to the inherent variations present in biological imaging data, particularly where data scarcity often limits the efficacy of conventional approaches.

Next, we sought to test the impact of the modifications we introduced in the CycleGAN-based approach on the generation of synthetic images of cells by comparing synthetic cell images generated by three different generator configurations: 1) UNET generator with L1 loss, 2) StyleUNET generator with L1 loss, and 3) StyleUNET generator with L1+VGG loss (Figure 3) and used the DeepSea model for the segmentation part. Synthetic images generated by the UNET model with L1 had an average Frechet Inception Distance (FID) score of 98, indicating a

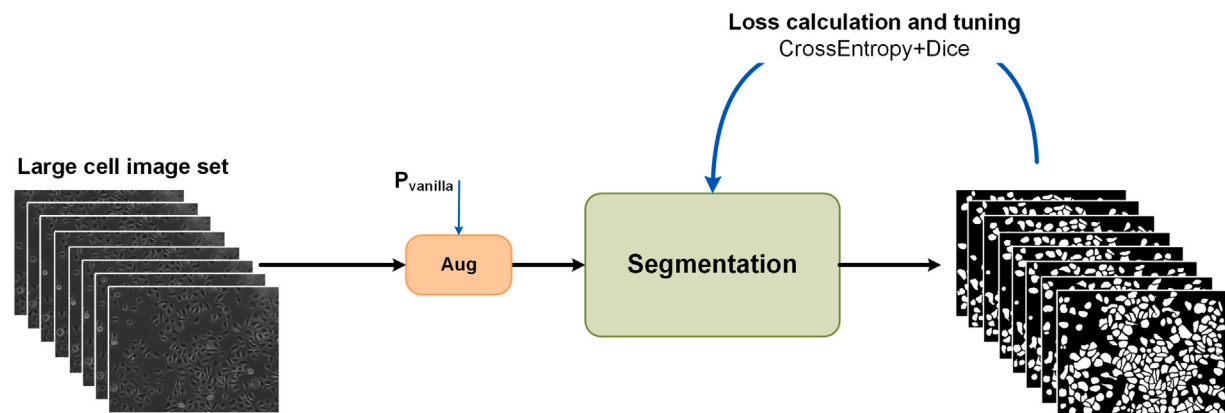
A



B



C

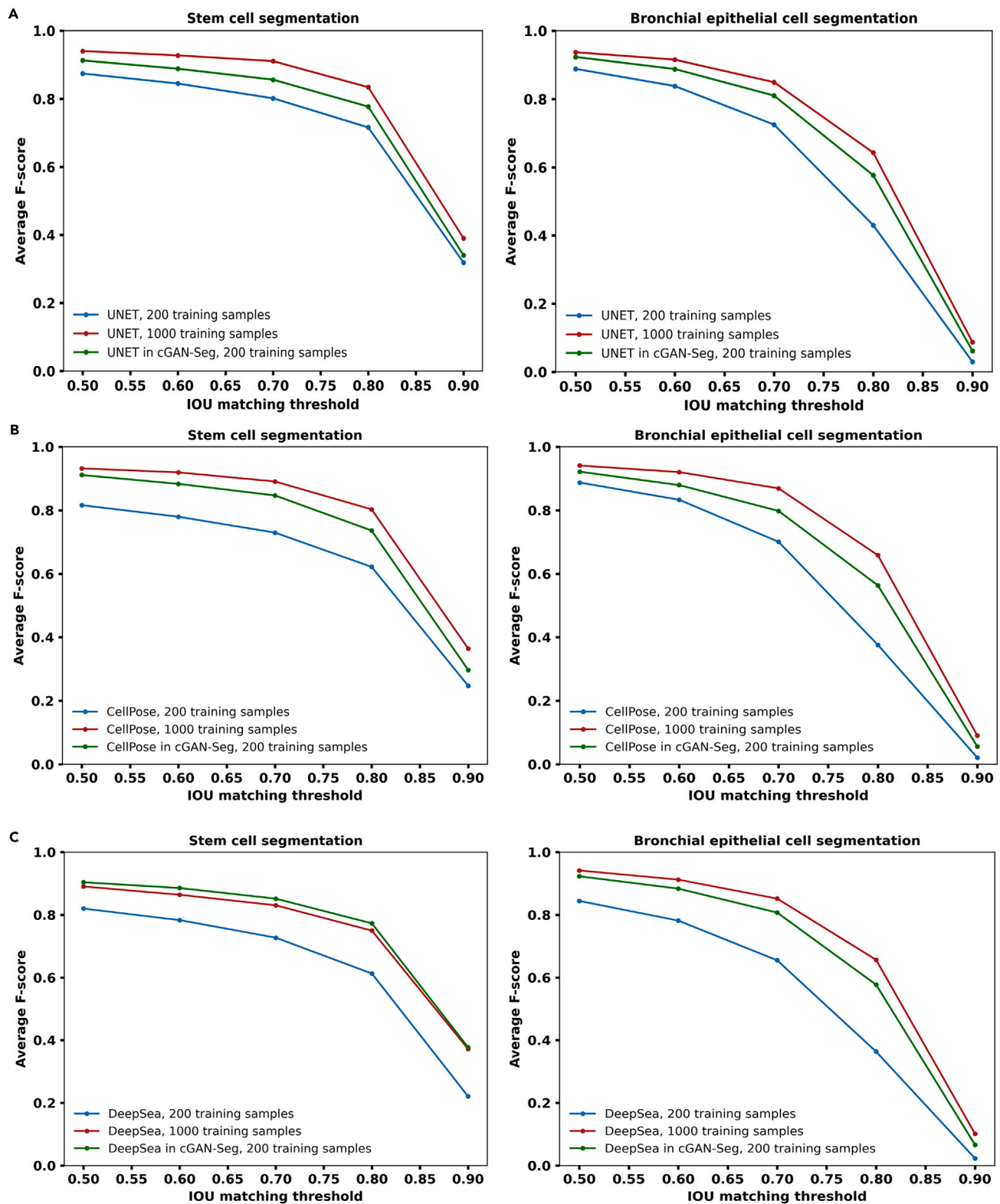


**Figure 1. Three different cell segmentation training scenarios**

(A) Training cGAN-Seg on a limited dataset of cell images (e.g., 200 training samples in this article) using a customized design of CycleGAN approach, termed cGAN-Seg, that incorporates features such as style injecting, modified PatchGan discriminator, and differentiable image augmentation.

(B) Conventional training of the segmentation model on a limited dataset of cell images.

(C) Conventional training of the segmentation model on a large dataset of cell images (e.g., 1000 training samples in this article).



**Figure 2. Comparative f-score performance of segmentation models across IoU thresholds**

F-Score performance across varied IoU thresholds for different segmentation models using cGAN-Seg training vs. conventional training on datasets of 200 and 1000 samples. A) UNET (top row), B) CellPose (middle row), and C) DeepSea (bottom row).



**Table 1. F-Score comparison for cGAN-Seg and conventional segmentation**

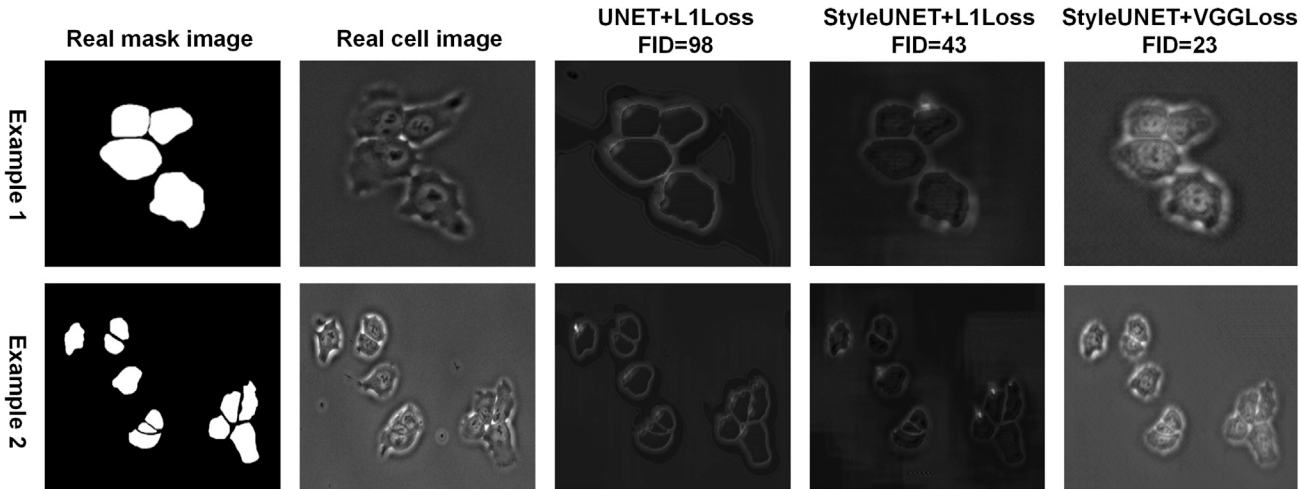
	A172 <sup>13</sup>	BT474 <sup>13</sup>	Huh7 <sup>13</sup>	SkBr3 <sup>13</sup>	C2C12 Muscle <sup>12</sup>	PSC <sup>23</sup>	U373 <sup>23</sup>
Conventional Seg: CellPose	0.51	0.53	0.47	0.89	0.79	0.90	0.95
Conventional Seg: 2D-UNET	0.52	0.53	0.49	0.90	0.79	0.92	0.96
cGAN-Seg Seg: CellPose	0.60	0.56	0.55	0.91	0.84	0.92	0.98
cGAN-Seg Seg: 2D-UNET	0.60	0.55	0.59	0.92	0.83	0.95	0.98

Comparative analysis of average f-scores at 0.5 IoU threshold: cGAN-Seg segmentation vs. conventional segmentation using limited datasets (200 training samples) across diverse cell types.

high dissimilarity between the synthetic and real cell images. The FID score dropped to 43 when we applied StyleUNET with L1 loss, resulting in increased similarities between synthetic images and real images. This improvement primarily stems from the enhanced capability of StyleUNET to capture and generate variations in style. Lastly, when we applied StyleUNET with the VGG perceptual loss, the synthetic images achieved a remarkably enhanced FID score of 23, reflecting a substantial increase in similarity to real cell images. The overall results of these modifications are synthetic images with detailed representations of features at subcellular levels. The enhanced similarity of synthetic images and lower FID score signifies the superiority of the perceptual loss function in preserving high-level details and morphological nuances, thereby leading to more realistic synthetic images.

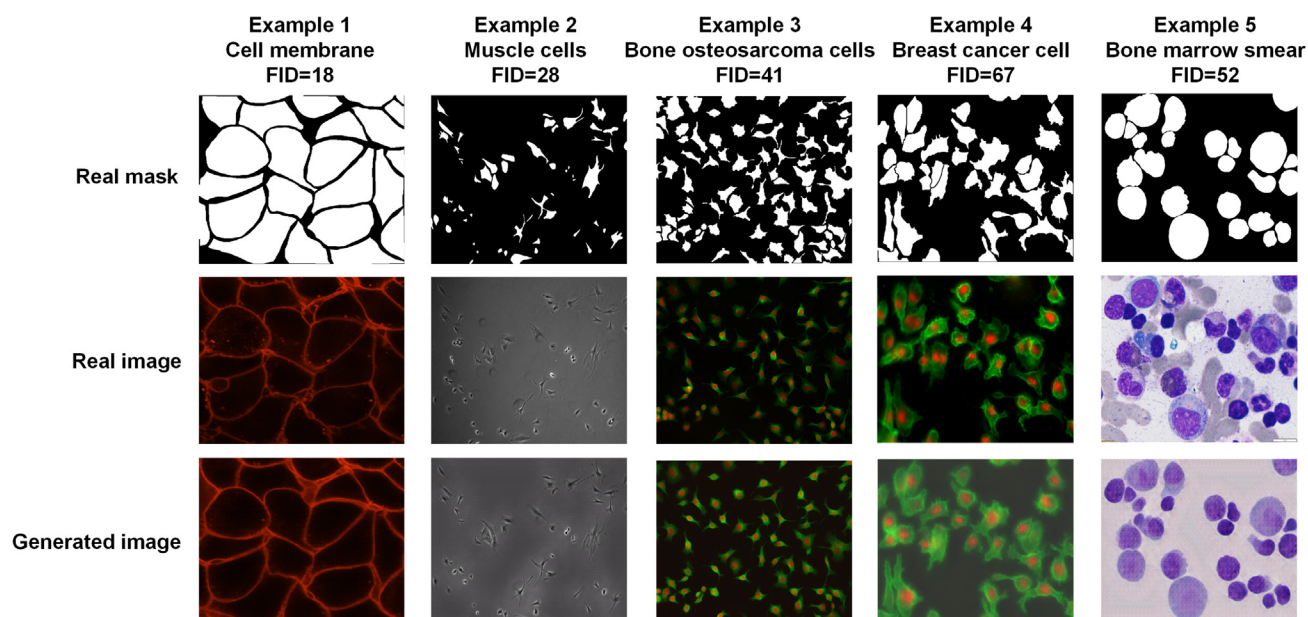
To demonstrate the extensibility of our proposed approach, we trained the cGAN-Seg model across various imaging modalities, diverse cell types, and subcellular organelles (from the DeepSea and CellPose datasets). Subsequently, the trained StyleUNET generator was employed to produce synthetic images that span these diverse conditions. While the limited number of available annotated images posed a challenge for a subset of samples, our StyleUNET generation model still generated realistic synthetic images similar to the real domain images and relatively overall low FID scores (Figure 4). We anticipate that with more extensive training data, the model's capacity for generating high-quality images would be significantly enhanced.

Most of our training data consisted of low- and mid-density cell images. We were curious to test if our generator could extrapolate its knowledge by generating new synthetic high-density images of cells that were not seen during training. This is particularly useful as the manual annotation of high-density images of cells can be very time-consuming and error-prone. To accomplish this, we designed an algorithm for generating synthetic high-density and colony-like cell masks (a relatively easy task) as input for the generator (in the test phase). As shown in Figure 5, our approach confirms the ability to extrapolate knowledge from low- and mid-density cell images, creating annotated images across any density level and magnification. This also includes the generation of colony-like cell formations (Example 3 in Figure 5) extending beyond the variations present in the original dataset. Such a capacity enables us to create synthetic cell images mimicking a broad



**Figure 3. Impact of style injection and vgg perceptual loss on cGAN-Seg image generation**

Two examples of comparing the effect of the style injecting technique and Vgg perceptual feature loss function on cGAN-Seg performance to generate images of DeepSea embryonic stem cells.



**Figure 4. Versatility of cGAN-Seg in generating realistic synthetic images across modalities**

cGAN-Seg style generator can generate synthetic images across multiple imaging modalities, cell types, and subcellular organelles (from DeepSea and CellPose datasets) similar to real images, as relatively low FID scores show.

range of real-world scenarios. The capacity of our model to extrapolate learned knowledge to unseen scenarios can provide a powerful tool to generalize this approach, aiming to develop segmentation models for a variety of cellular imaging modalities.

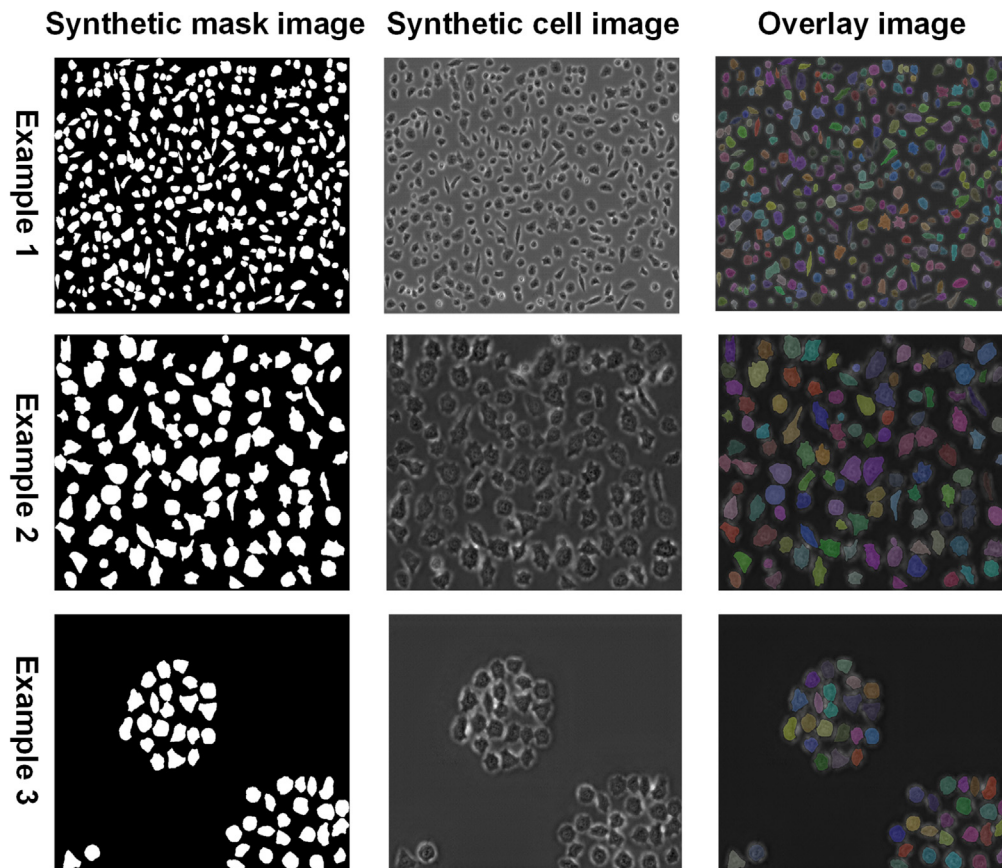
To further assess the capability of the StyleUNET generation model in enhancing the segmentation of more challenging, high-density cell images, we used our synthetic high-density cell images (showcased in Figure 5) to conventionally train a segmentation model. To validate the model evaluation process, we categorized the test set images into two groups based on their complexity: “easy” samples, which include isolated cells or cells in non-touching colonies, and “hard” samples, characterized by cells in close contact or within touching colonies. In this experiment, we employed the DeepSea model for the segmentation and focused on the DeepSea stem cell test set, where we found enough easy and hard samples for this specific analysis. As presented in Table 2, integrating hard, colony-like, and high-density synthetic images into the training process notably enhances segmentation performance on “hard” samples to a greater extent compared to the “easy” samples, underscoring the effectiveness of StyleUNET generation model outputs in tackling complex segmentation scenarios.

## DISCUSSION

A large and diverse annotated dataset of images is key to the successful development of deep learning models that can perform across a variety of real-world images. Currently, a harmonized large and diverse dataset of microscopy images is not available to train new deep-learning models because the annotation of microscopy images is a tedious and time-consuming task. Our study provides a solution to this problem by proposing a method for training cell segmentation models using a CycleGAN approach that we termed cGAN-Seg to address the critical issue of limited annotated data in cellular imaging. cGAN-Seg harnesses the potential of GANs to generate a diverse set of synthetic realistic cell images, enhance the diversity of the available training datasets without manual annotation, and improve the overall performance of the segmentation models with a limited annotated dataset. Importantly, we showed that cGAN-Seg allows for the extrapolation of knowledge by model by generating synthetic images that the model has not been exposed to during the training.

We made several modifications to the original CycleGAN architecture to build the cGAN-Seg model and apply the microscopy images. First, a style generation path was integrated into the synthetic image generator to boost variation in synthetic images. Second, a linear attention mechanism was incorporated into the PatchGAN discriminator-based architecture to enhance its differentiation capabilities and synthetic image quality. Third, differentiable image augmentation was introduced during the training phase to further diversify image generation and reduce the risk of overfitting. Fourth, instead of the L1 loss function conventionally used in the CycleGAN, we employed a combination of cross-entropy (CE) and dice losses for the segmentation, improving the handling of multi-class classification and imbalanced datasets. Finally, as a critical modification, we replaced the L1 loss function in the generator with a VGG perceptual loss function to promote the retention of more high-level features and nuances in the generated synthetic images, leading to enhanced similarity between real images of cells and synthetic images. These enhancements collectively improved the diversity and quality of synthetic cell images, resulting in a more diverse and generalized segmentation model trained with various microscopy imaging styles and conditions.





**Figure 5. Generating high-density cell images with cGAN-Seg**

Three examples of producing colony-like and high-density cell images using synthetic high-density mask images as input for cGAN-Seg style generator.

Our experimental results show that the proposed cGAN-Seg approach provides a straightforward solution for the paucity of annotated microscopy data for training deep-learning models. Experimental results showed that the performance of segmentation models trained using our cGAN-Seg method improved the segmentation scores across different cell types of different datasets. Notably, this enhancement was observed irrespective of the scarcity of annotated cell image datasets, illustrating the potential of our approach in effectively addressing this prevalent issue in biomedical imaging. Implementation of Style injecting in our UNET generator significantly improved the quality of synthetic images, reflected by an FID score reduction from 98 to 43. A further enhancement was achieved by adding VGG perceptual loss to the conventional L1 loss function, resulting in an FID score of 23. We also validated the model's versatility across multiple imaging modalities, cell types, and sub-cellular organelles. Besides, we further validated the trained cGAN-Seg ability to extrapolate knowledge from low- and mid-density cell images, creating annotated images across different density levels and magnifications, even those absent in the original training dataset. We extended the utility of the cGAN-Seg model by incorporating the generated synthetic high-density cell images into the segmentation fine-tuning process. This addition enhanced the model's segmentation capabilities, especially for complex, densely populated cell structures.

In our experiments, we noticed that the impact of the cGAN-Seg training approach varied between different cell-type images. The dataset comprising less complexity and lower diversity showed a lower improvement compared to the more complex dataset. This can be attributed to the inherent simplicity of the data, which likely enabled the model to learn necessary patterns without the need for additional augmented examples. Conversely, for complex datasets with more inherent variability, cGAN-Seg proved more beneficial by providing diverse synthetic image examples during the training, thus enhancing the model's ability to generalize. We believe that the exact dynamics depend on the specific dataset, model, and augmentation techniques used, underscoring the need for task-specific experimentation and validation.

Our CycleGAN-based method opens up new possibilities for training deep-learning models for microscopy applications by offering a solution to the challenge of limited annotated cell image datasets. This study illustrates how generative deep learning methods such as GANs can be utilized to address data limitations in microscopy, thereby pushing the boundaries of what is possible in the field of biomedical imaging. It is important to mention that while our approach has shown promising results, there is room for further improvement and experimentation, including exploring different GAN architectures and further refinement of the augmentation techniques. The generated synthetic images can also be made more diverse and realistic through additional modifications in the GAN training process.

**Table 2. Enhancing segmentation model training with synthetic high-density images**

	200 training images	200 training images+700 synthetic images
Easy samples	0.86	0.92
Hard samples	0.77	0.86

Average f-score comparison before and after adding synthetic high-density cell images (generated by the cGAN-Seg style generator) into the conventional training process of the DeepSea segmentation model.

### Limitation of the study

While our cGAN-Seg approach enhances the training of cell segmentation models in the context of limited annotated datasets, there are a few limitations to consider. Firstly, the quality and diversity of the generated synthetic images, although improved, may not fully capture the complexity of all possible cell types and conditions found in real-world microscopy images. This could lead to model performance that is not uniformly enhanced across different imaging modalities or biological contexts. Secondly, the integration of style generation paths and the use of VGG perceptual loss in our modified CycleGAN architecture can introduce additional complexity and computational demands, which may limit the accessibility of our approach to researchers with fewer computational resources.

### STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- **KEY RESOURCES TABLE**
- **RESOURCE AVAILABILITY**
  - Lead contact
  - Materials availability
  - Data and code availability
- **METHOD DETAILS**
  - Datasets
  - CycleGAN-based approach overall view
  - Augmentation functions
  - Segmentation models
  - Style generation model
  - Discriminator models
  - Evaluation metrics
  - Loss functions
- **QUANTIFICATION AND STATISTICAL ANALYSIS**

### SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.isci.2024.109740>.

### ACKNOWLEDGMENTS

We would like to thank the members of the Shariati lab for their feedback on this work. This work was supported by the NIGMS/NIH through a Pathway to Independence Award K99GM126027/R00GM126027 and Maximizing Investigator Award (R35GM147395), a start-up package from the University of California, Santa Cruz (S.A.S). B.T is supported by the National Institutes of Health (NIH) under award number K12GM139185 and the Institute for the Biology of Stem Cells (IBSC) at UC Santa Cruz. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH or the IBSC.

### AUTHOR CONTRIBUTIONS

A.Z.K and S.A.S were pivotal in conceptualizing the project and played a significant role in article preparation. Both A.Z.K and N.M developed the related Python scripts and prepared the article materials. B.T helped in dataset collection and article review processes.

### DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: November 1, 2023

Revised: February 20, 2024

Accepted: April 10, 2024

Published: April 12, 2024

## REFERENCES

- Karras, T., Laine, S., and Aila, T. (2019). A Style-Based Generator Architecture for Generative Adversarial Networks, pp. 4396–4405.
- Ledig, C., Theis, L., Huszár, F., Caballero, J., Cunningham, A., Acosta, A., Aitken, A., Tejani, A., Totz, J., Wang, Z., and Shi, W. (2017). Photo-Realistic Single Image Super-resolution Using a Generative Adversarial Network, pp. 105–114.
- Arjovsky, M., Chintala, S., and Bottou, L. (2017). Wasserstein Generative Adversarial Networks. In *Proceedings of the 34th International Conference on Machine Learning*, P. Doina and T. Yee Whye, eds. (PMLR).
- Goodfellow, I.J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., and Bengio, Y. (2014). Generative Adversarial Networks. Preprint at arXiv. <https://doi.org/10.48550/arXiv.1406.2661>.
- Mao, X., Li, Q., Xie, H., Lau, R.Y.K., Wang, Z., and Smolley, S.P. (2017). Least Squares Generative Adversarial Networks. 22-29 Oct. 2017, pp. 2813–2821.
- Lotter, W., Kreiman, G., and Cox, D. (2016). Deep Predictive Coding Networks for Video Prediction and Unsupervised Learning. Preprint at arXiv. <https://doi.org/10.48550/arXiv.1605.08104>.
- Thirumagal, E., and Saruladha, K. (2021). Chapter 2 - GAN models in natural language processing and image translation. In *Generative Adversarial Networks for Image-to-Image Translation*, A. Solanki, A. Nayyar, and M. Naved, eds. (Academic Press), pp. 17–57. <https://doi.org/10.1016/B978-0-12-823519-5.00001-4>.
- Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., Xiao, T., Whitehead, S., Berg, A.C., Lo, W.-Y., et al. (2023). Segment Anything. Preprint at arXiv. <https://doi.org/10.48550/arXiv.2304.02643>.
- Padovani, F., Mairhörmann, B., Falter-Braun, P., Lengefeld, J., and Schmoller, K.M. (2021). Cell-ACDC: a user-friendly toolset embedding state-of-the-art neural networks for segmentation, tracking and cell cycle annotations of live-cell imaging data. Preprint at bioRxiv. <https://doi.org/10.1101/2021.09.28.462199>.
- Schmidt, U., Weigert, M., Broaddus, C., and Myers, G. (2018). In *Cell Detection with Star-Convex Polygons*. held in Cham, A.F. Frangi, J.A. Schnabel, C. Davatzikos, C. Alberola-López, and G. Fichtinger, eds. (Springer International Publishing), pp. 265–273.
- Stringer, C., Wang, T., Michaelos, M., and Pachitariu, M. (2021). Cellpose: a generalist algorithm for cellular segmentation. *Nat. Methods* 18, 100–106. <https://doi.org/10.1038/s41592-020-01018-x>.
- Zargari, A., Lodewijk, G.A., Mashhadi, N., Cook, N., Neudorf, C.W., Araghbidikashani, K., Hays, R., Kozuki, S., Rubio, S., Hrabeta-Robinson, E., et al. (2023). DeepSea is an efficient deep-learning model for single-cell segmentation and tracking in time-lapse microscopy. *Cell Rep. Methods* 3, 100500. <https://doi.org/10.1016/j.crmeth.2023.100500>.
- Edlund, C., Jackson, T.R., Khalid, N., Bevan, N., Dale, T., Dengel, A., Ahmed, S., Trygg, J., and Sjögren, R. (2021). LIVECell—A large-scale dataset for label-free live cell segmentation. *Nat. Methods* 18, 1038–1045. <https://doi.org/10.1038/s41592-021-01249-6>.
- Archit, A., Nair, S., Khalid, N., Hilt, P., Rajashekar, V., Freitag, M., Gupta, S., Dengel, A., Ahmed, S., and Pape, C. (2023). Segment Anything for Microscopy. Preprint at bioRxiv 2023. <https://doi.org/10.1101/2023.08.21.554208>.
- Zhu, J.Y., Park, T., Isola, P., and Efros, A.A. (2017). Unpaired Image-To-Image Translation Using Cycle-Consistent Adversarial Networks, pp. 2242–2251. 22-29.
- Karras, T., Laine, S., Aittala, M., Hellsten, J., Lehtinen, J., and Aila, T. (2019). Analyzing and Improving the Image Quality of StyleGAN. Preprint at arXiv. <https://doi.org/10.48550/arXiv.1912.04958>.
- Ronneberger, O., Fischer, P., and Brox, T. (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation. Preprint at arXiv. <https://doi.org/10.48550/arXiv.1505.04597>.
- Isola, P., Zhu, J.Y., Zhou, T., and Efros, A.A. (2017). Image-to-Image Translation with Conditional Adversarial Networks, pp. 5967–5976.
- Jetley, S., Lord, N.A., Lee, N., and Torr, P.H.S. (2018). Learn To Pay Attention. Preprint at arXiv. <https://doi.org/10.48550/arXiv.1804.02391>.
- Zhao, S., Liu, Z., Lin, J., Zhu, J.-Y., and Han, S. (2020). Differentiable Augmentation for Data-Efficient GAN Training. Preprint at arXiv. <https://doi.org/10.48550/arXiv.2006.10738>.
- Creswell, A., White, T., Dumoulin, V., Arulkumaran, K., Sengupta, B., and Bharath, A.A. (2018). Generative Adversarial Networks: An Overview. *IEEE Signal Process. Mag.* 35, 53–65. <https://doi.org/10.1109/MSP.2017.2765202>.
- Johnson, J., Alahi, A., and Fei-Fei, L. (2016). Perceptual Losses for Real-Time Style Transfer and Super-Resolution. Preprint at arXiv. <https://doi.org/10.48550/arXiv.1603.08155>.
- Maška, M., Ulman, V., Delgado-Rodríguez, P., Gómez-de-Mariscal, E., Nečasová, T., Guerrero Peña, F.A., Ren, T.I., Meyerowitz, E.M., Scherr, T., Löffler, K., et al. (2023). The Cell Tracking Challenge: 10 years of objective benchmarking. *Nat. Methods* 20, 1010–1020. <https://doi.org/10.1038/s41592-023-01879-y>.
- Lim, S., Kim, I., Kim, T., Kim, C., and Kim, S. (2019). Fast AutoAugment. Preprint at arXiv. <https://doi.org/10.48550/arXiv.1905.00397>.
- Yang, S., Xiao, W., Zhang, M., Guo, S., Zhao, J., and Shen, F. (2022). Image Data Augmentation for Deep Learning: A Survey. Preprint at arXiv. <https://doi.org/10.48550/arXiv.2204.08610>.
- Müller, D., Soto-Rey, I., and Kramer, F. (2022). Towards a guideline for evaluation metrics in medical image segmentation. Preprint at arXiv. <https://doi.org/10.48550/arXiv.2202.05273>.
- Heusel, M., Ramsauer, H., Unterthiner, T., Nessler, B., and Hochreiter, S. (2017). GANs Trained by a Two Time-Scale Update Rule Converge to a Local Nash Equilibrium. Preprint at arXiv. <https://doi.org/10.48550/arXiv.1706.08500>.
- Dosovitskiy, A., and Brox, T. (2016). Generating Images with Perceptual Similarity Metrics based on Deep Networks. Preprint at arXiv. <https://doi.org/10.48550/arXiv.1602.02644>.

## STAR★METHODS

### KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Software and algorithms		
CellPose	Stringer et al. <sup>11</sup>	<a href="https://www.cellpose.org/">https://www.cellpose.org/</a>
2D-UNET	Ronneberger et al. <sup>17</sup>	N/A
DeepSea	Zargari et al. <sup>12</sup>	<a href="https://deepseas.org/">https://deepseas.org/</a>
cGAN-Seg	This paper	<a href="https://github.com/abzargar/cGAN-Seg">https://github.com/abzargar/cGAN-Seg</a> Zenodo <a href="https://doi.org/10.5281/zenodo.10931977">https://doi.org/10.5281/zenodo.10931977</a>

### RESOURCE AVAILABILITY

#### Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, S. Ali Shariati ([alish@ucsc.edu](mailto:alish@ucsc.edu)).

#### Materials availability

All unique/stable materials generated in this study are available from the [lead contact](#) upon reasonable request with a completed Materials Transfer Agreement.

#### Data and code availability

- The image datasets utilized in our study are publicly available via a link on our GitHub repository: <https://github.com/abzargar/cGAN-Seg>, facilitating easy access for replication and further research efforts.
- The Python scripts encompassing the methodologies we developed are publicly accessible for download at our GitHub repository: <https://github.com/abzargar/cGAN-Seg>.
- Any additional information required to reanalyze the data reported in this work paper is available from the [lead contact](#) upon request.

### METHOD DETAILS

#### Datasets

In our study, we utilized four distinct training datasets, each representing different modalities and cell types, to ensure the robustness and generalizability of our proposed segmentation model across a diverse range of biological contexts. They include 1) our recently published annotated dataset of phase-contrast images of the DeepSea,<sup>12</sup> which is a large collection of accurately annotated phase-contrast time-lapse microscopy images of three cell types of Mouse Embryonic Stem Cells, Bronchial epithelial cells, and Mouse C2C12 Muscle Progenitor Cells, 2) LiveCell dataset<sup>13</sup> which is a diverse collection of annotated microscopy images. It covers various cell types such as A172, BT474, Huh7, and SkBr3 across multiple experimental conditions, 3) Cell Tracking Challenge dataset,<sup>23</sup> which is a dataset repository consisting of 2D and 3D time-lapse sequences of fluorescent images of different cell types such as PSC and U373 cells, 4) CellPose dataset<sup>11</sup> which is a curated collection encompassing a wide variety of annotated images from different cell types, tissues, and organisms. We used 15% of the dataset samples for testing all training scenarios. To ensure a rigorous evaluation of our model, we allocated 15% of the samples from each dataset exclusively for testing across all training scenarios.

#### CycleGAN-based approach overall view

In our proposed cGAN-Seg method (Figure 1A), in addition to the conventional augmentation pipeline, we employed an approach to train the cell image segmentation model using a modified CycleGAN process.<sup>15</sup> In comparison with conventional augmentation techniques, the CycleGAN process enhances the diversity of training cell images for the segmentation model by generating synthetic images of cells and their corresponding masks that are almost indistinguishable from real images. The CycleGAN architecture is unique in that every step in the training process encompasses two mapping paths: forward consistency and backward consistency (Equation 1). In the forward consistency path, the model learns to translate an image from domain A (mask image) to domain B (cell image). Subsequently, it attempts to translate this new cell image back to domain A, aiming to reconstruct the original mask image. The backward consistency path mirrors this process, starting with domain B, translating to domain A, and then back to B. The essence of this approach lies in its ability to maintain the integrity of the original images throughout the translation and back-translation processes, ensuring that crucial information is not lost. This dual-path mechanism has the potential to introduce a greater degree of diversity to the training data. Within this architecture, the generator plays a

pivotal role. It is capable of creating new synthetic cell images that might exhibit a significantly different distribution pattern from the original, real training samples. As a result, our segmentation model can train on a blend of these artificially generated images and the augmented real images. This hybrid training approach not only diversifies the data pool but also helps the model adapt to a broader spectrum of cell images. The addition of synthetic images simulates a wider array of scenarios that the model may encounter, thus improving its robustness and predictive power when faced with unfamiliar data.

$$\left\{ \begin{array}{l} \text{Forward Consistency Path : Real mask image} \xrightarrow{\text{Generation}} \text{Fake cell image} \xrightarrow{\text{Segmentation}} \text{Fake mask image} \\ \text{Backward Consistency Path : Real cell image} \xrightarrow{\text{Segmentation}} \text{Fake mask image} \xrightarrow{\text{Generation}} \text{Fake cell image} \end{array} \right. \quad (\text{Equation 1})$$

### Augmentation functions

Image augmentation techniques play a critical role in expanding the diversity of training datasets, thereby improving model generalization and robustness.<sup>24,25</sup> In the training process of our deep learning models, we applied some mostly used conventional image augmentation functions to every single cell image with the probability of  $p_{\text{vanilla}}$ , including random histogram equalization, random crop, random sharpness adjustment, random brightness adjustment, random contrast adjustment, random horizontal flip, random vertical flip, random saturation, adding random gaussian noise, and adding random gaussian blur as shown in Figure S1A. For the binary mask images, we only used the applicable random crop, horizontal flip, and vertical flip functions. The training algorithm executes a sequence of the provided augmentation functions for each cell and mask image pair with a pre-defined probability value ' $p_{\text{vanilla}}$ '. In the requested augmentation pipeline, each function is randomly chosen with a consistent probability of 50% and is also applied in a randomized sequence.

However, when it comes to training Generative Adversarial Networks (GANs), especially models like CycleGAN that learn mappings between different image domains, conventional augmentation might not be sufficient for enhancing the diversity of generated images. This is where differentiable augmentation becomes valuable. Differentiable augmentation applies the same random augmentations to both real and fake samples in a way that is differentiable with respect to the model parameters.<sup>20</sup> This approach encourages the discriminator to less memorize the exact training samples, thus causing the generator to produce more diverse images, thereby improving the overall image generation performance. Furthermore, differentiable augmentation can mitigate overfitting and improve training stability, making it particularly beneficial for GANs trained with limited data. In this project, we used five different differentiable augmentation functions: random contrast, random brightness, random cutout, random translation, and random saturation (Figure S1B). The decision to perform an augmentation is dictated by the probability variable ' $p_{\text{diff}}$ '. In an attempt to ensure fair representation and randomness, each of the differentiable augmentation functions is executed in a randomized sequence, with each having an equal 50% probability of selection. This approach not only diversifies the images but also ensures that the model remains adaptable to any new form of data it might encounter in the future, thus improving its resilience and overall effectiveness.

### Segmentation models

For the segmentation tasks, we employed and compared our recently published DeepSea baseline architecture<sup>12</sup> along with two widely used segmentation models: the 2D-UNET<sup>17</sup> and CellPose<sup>11</sup> models. The 2D-UNET model has become synonymous with high-performance image segmentation across various biomedical applications. Its design is characterized by a symmetric encoder-decoder structure that efficiently captures context and enables precise localization. This architecture facilitates the learning of rich feature representations from limited training data, making it particularly suitable for medical imaging tasks where annotated samples are scarce. The strength of 2D-UNET lies in its ability to handle a wide range of cell types and imaging conditions, thanks to its deep convolutional layers and skip connections that preserve spatial information across the network. DeepSea architecture is also an efficient scaled-down version of the 2D-UNET mode. To simplify the task, we chose not to incorporate the layer representing touching cells, as this would necessitate custom touching cell masks (alongside cell body masks) and additional loss functions. The CellPose model also represents a significant leap forward in the segmentation of complex cell images. It is designed around the concept of predicting cell 'poses'—spatial arrangements that are invariant to cell shape and size—allowing it to segment cells in a highly generalized manner. This model leverages a powerful neural network trained on a diverse dataset, enabling it to accurately segment cells across different experiments without the need for retraining. The CellPose model's robustness and adaptability stem from its novel use of flow fields, which guide the segmentation process and ensure high precision across varying biological contexts. We would like to mention that to reduce complexity and focus on the essential aspects of segmentation, we excluded CellPose's flow fields from the cGAN-Seg approach.

### Style generation model

In our proposed cGAN-Seg architecture, the generator is responsible for generating synthetic cell images. It employs a 2D-UNET architecture,<sup>17</sup> as shown in Figure S2. UNET is renowned for its effectiveness in biomedical image segmentation due to its unique architecture, which consists of a contracting path to capture context and a symmetric expanding path that enables precise localization. However, we have taken this a step further by incorporating a style decoding path into the decoder part of the UNET architecture, an idea inspired by the StyleGAN2 model.<sup>16</sup>

This fusion of concepts from StyleGAN2 and UNET brings about the prospect of generating better synthetic images. The style decoding network is designed to control the stylistic aspects of the generated images, thereby allowing the model to create more diverse and

potentially higher-quality synthetic cell images. This combination of architectures seeks to maximize the strengths of both models - the segmentation prowess of UNET and the sophisticated generative capacity of StyleGAN2. This integration could potentially yield a more powerful generator model for synthetic cell image creation, thereby enhancing the overall performance of our cGAN-Seg model.

### Discriminator models

In our proposed method, we applied a modified version of the PatchGAN baseline architecture<sup>18</sup> for our discriminators, integrating a layer of residual linear attention, as shown in [Figure S3](#). The PatchGAN architecture, known for its effectiveness in examining both global and local image features, has demonstrated impressive performance in diverse GAN applications. The architecture operates by creating multiple 'paths' with different receptive field sizes, enabling the model to scrutinize image details at various scales.

However, we sought to improve the discriminator's ability to focus on critical features by incorporating an additional layer of residual linear attention.<sup>19</sup> This is an approach to attention mechanisms that makes use of a linear combination of input features and learned attention maps, thereby enabling the model to weigh different regions of the input differently. As a result, the model can focus on more critical parts of the image, thereby enhancing its ability to discriminate real images from synthetic ones accurately. By modifying the PatchGAN discriminator, we aimed to improve the model's focus on salient image features, thus boosting its ability to accurately distinguish between real and generated images.

### Evaluation metrics

In the testing phase, we leveraged the Intersection over Union (IoU) metric, also known as the Jaccard Index, which ranges from 0 to 1, to evaluate the alignment between the segmentation model's predictions and the manually annotated ground truth masks.<sup>26</sup> For each test image, we designated each detected cell body as a True Positive (TP) prediction if its IoU index exceeded a predetermined threshold value, indicating a valid match to the ground truth. Conversely, any ground truth cell body masks that failed to find a valid match were classified as False Negatives (FN), and any predictions lacking corresponding ground truth masks were labeled as False Positives (FP), representing non-cell entities. Subsequently, we calculated the Precision, Recall, and F-score for each image in the test set using [Equations 2, 3, and 4](#).

$$Precision = \frac{TP}{TP+FP} \quad (\text{Equation 2})$$

$$Recall = \frac{TP}{TP+FN} \quad (\text{Equation 3})$$

$$Fscore = 2 \times \frac{Precision \times Recall}{Precision+Recall} \quad (\text{Equation 4})$$

We also evaluated our segmentation models using Dice Score. The Dice score, also known as the Dice Similarity Coefficient, is a common metric used in the field of image segmentation, including cell image segmentation, to measure the similarity between two samples and can be measured by [Equation 5](#), X represents the set of pixels in the predicted segmentation and Y represents the set of pixels in the ground truth segmentation.

$$Dice\ Score = 2 \times \frac{|X \cap Y|}{|X| + |Y|} \quad (\text{Equation 5})$$

Frechet Inception Distance (FID) is also a widely used metric to evaluate the quality of images generated by GAN models.<sup>27</sup> It quantifies the dissimilarity between the distributions of generated and real images in the feature space of a pre-trained Inception network. Lower FID scores represent higher-quality synthetic images that more closely resemble the distribution of real images. By assessing differences in both mean and covariance of the features, FID provides a more comprehensive evaluation of image quality and diversity, making it a good choice for evaluating the performance of our modified generator.

### Loss functions

In the training phase of our cGAN-Seg model, we employed a series of loss functions to effectively optimize the performance of both the generator and the segmentation model. These loss functions are specifically tailored to address the unique challenges presented by the task of generating high-quality synthetic images and accurately segmenting cell structures.

Two fundamental loss functions utilized in our model are identity loss and reconstruction loss. Identity loss ensures that an image translated to its own domain remains unchanged, which encourages the generator to preserve color and texture composition between the input and output.<sup>15</sup> Reconstruction loss, on the other hand, is used to maintain cycle consistency, ensuring that an image translated from one domain to another can be accurately translated back to its original form. These losses are critical to ensure that the model not only learns the correct mappings between the domains but also produces images that are consistent with the original data distribution.

For the generator, we utilized the VGG perceptual feature loss function<sup>22,28</sup> for both identity and reconstruction loss ([Equations 8 and 10](#)). The VGG loss function is a high-level feature extraction loss that helps preserve the perceptual and semantic understanding of the images. It is



a concept based on a deep convolutional neural network (CNN), like VGG, that has been pre-trained on a large dataset for an image classification task.

For the segmentation model, we combined Cross-Entropy (CE) loss and Dice loss for both identity and reconstruction loss (Equations 9 and 11). Cross-Entropy loss is a popular choice for multi-class classification problems, calculating the dissimilarity between the predicted probability distribution and the ground truth distribution. The Dice loss, on the other hand, is specifically designed for handling imbalanced datasets and is extensively used in medical image segmentation tasks due to its efficiency in dealing with small objects and imbalanced classes. By using these two loss functions in tandem, we enhance the performance of our segmentation model, ensuring it can effectively handle the challenges of cell image segmentation.

The discriminators in our model were optimized using the Mean Squared Error (MSE) loss as an adversarial loss. This loss function encourages the discriminators to distinguish between real and fake images by minimizing the average squared differences between the predicted and actual values.

Each of these loss functions is assigned a specific weight in order to balance their contributions during the optimization process (Equations 16 and 17). By integrating these diverse loss functions and carefully selecting their weights, we can effectively train our cGAN-Seg model, ensuring both the production of diverse, high-quality synthetic cell images and the accurate segmentation of cell structures.

$$D1\_G\_L = MSE(1, D1(DiffAug(Gen(real\_mask)))) \quad (\text{Equation 6})$$

$$D2\_S\_L = MSE(1, D2(Seg(real\_img))) \quad (\text{Equation 7})$$

$$Rec\_G\_L = L1(real\_img, Gen(Seg(real\_img))) + VGG(real\_img, Gen(Seg(real\_img))) \quad (\text{Equation 8})$$

$$Rec\_S\_L = CE(real\_mask, Seg(Gen(real\_mask))) + Dice(real\_mask, Seg(Gen(real\_mask))) \quad (\text{Equation 9})$$

$$Id\_G\_L = L1(real\_img, Gen(real\_mask)) + VGG(real\_img, Gen(real\_mask)) \quad (\text{Equation 10})$$

$$Id\_S\_L = CE(real\_mask, Seg(real\_img)) + Dice(real\_mask, Seg(real\_img)) \quad (\text{Equation 11})$$

$$Real\_D1\_L = MSE(1, D1(DiffAug(real\_img))) \quad (\text{Equation 12})$$

$$Real\_D2\_L = MSE(1, D2(real\_mask)) \quad (\text{Equation 13})$$

$$Fake\_D1\_L = MSE(0, D1(DiffAug(Gen(real\_mask)))) \quad (\text{Equation 14})$$

$$Fake\_D2\_L = MSE(0, D2(Seg(real\_img))) \quad (\text{Equation 15})$$

$$Total\_Gen\_loss = D1\_G\_L + D2\_S\_L + 100 \times Rec\_G\_L + 100 \times Rec\_S\_L + 50 \times Id\_G\_L + 50 \times Id\_S\_L \quad (\text{Equation 16})$$

$$Total\_D\_loss = 0.5 \times Real\_D1\_L + 0.5 \times Real\_D2\_L + 0.5 \times Fake\_D1\_L + 0.5 \times Fake\_D2\_L \quad (\text{Equation 17})$$

## QUANTIFICATION AND STATISTICAL ANALYSIS

To verify the reproducibility of our findings, we repeated all training and testing experiments using the cross-validation method. We selected five random subsets for training, validation, and testing from the generated dataset and reported the average performance metrics.