HAL open science

A deep learning segmentation strategy that minimizes the amount of manually annotated images

Thierry Pécot, Alexander Alekseyenko, Kristin Wallace

▶ **To cite this version:**

Thierry Pécot, Alexander Alekseyenko, Kristin Wallace. A deep learning segmentation strategy that minimizes the amount of manually annotated images. F1000Research, 2022, 10, pp.256. 10.12688/f1000research.52026.2 . hal-03594564v2

HAL Id: hal-03594564

https://hal.science/hal-03594564v2

Submitted on 8 Dec 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Distributed under a Creative Commons Attribution 4.0 International License

Check for updates

METHOD ARTICLE

REVISED A deep learning segmentation strategy that minimizes the amount of manually annotated images [version 2; peer review: 2 approved]

Thierry Pécot[1,2], Alexander Alekseyenko[3], Kristin Wallace[4]

[1]Department of Biochemistry and Molecular Biology, Hollings Cancer Center, Medical University of South Carolina, Charleston, SC, 29407, USA
[2]Rennes 1 University, SFR Biosit (UMS 3480 - US 018), F-35000 Rennes, France
[3]Departments of Public Health Sciences and Oral Health Sciences, Biomedical Informatics Center, Medical University of South Carolina, Charleston, SC, 29407, USA
[4]Department of Public Health Sciences, Medical University of South Carolina, Charleston, SC, 29407, USA

v2 **First published:** 30 Mar 2021, **10**:256
https://doi.org/10.12688/f1000research.52026.1
Latest published: 17 Jan 2022, **10**:256
https://doi.org/10.12688/f1000research.52026.2

Abstract
Deep learning has revolutionized the automatic processing of images. While deep convolutional neural networks have demonstrated astonishing segmentation results for many biological objects acquired with microscopy, this technology's good performance relies on large training datasets. In this paper, we present a strategy to minimize the amount of time spent in manually annotating images for segmentation. It involves using an efficient and open source annotation tool, the artificial increase of the training dataset with data augmentation, the creation of an artificial dataset with a conditional generative adversarial network and the combination of semantic and instance segmentations. We evaluate the impact of each of these approaches for the segmentation of nuclei in 2D widefield images of human precancerous polyp biopsies in order to define an optimal strategy.

Keywords
Deep learning, image annotation, semantic and instance segmentations, conditional GANs, nuclei segmentation

Open Peer Review

Approval Status ✓ ✓

	1	2
version 2 (revision) 17 Jan 2022	✓ view	✓ view
version 1 30 Mar 2021	? view	? view

1. **Alice Lucas**, Broad Institute, Cambridge, USA

2. **Romain F Laine**, University College London, London, UK

Any reports and responses or comments on the article can be found at the end of the article.

This article is included in the NEUBIAS - the Bioimage Analysts Network gateway.

This article is included in the Artificial Intelligence and Machine Learning gateway.

This article is included in the Bioinformatics gateway.

Corresponding author: Thierry Pécot (thierry.pecot@univ-rennes1.fr)

Author roles: Pécot T: Conceptualization, Data Curation, Formal Analysis, Methodology, Software, Validation, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing; **Alekseyenko A**: Writing – Review & Editing; **Wallace K**: Funding Acquisition, Writing – Review & Editing

Competing interests: No competing interests were disclosed.

Grant information: This work was funded by a Chan Zuckerberg Initiative DAF grant to T.P. (2019-198009).
The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Copyright: © 2022 Pécot T *et al*. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

How to cite this article: Pécot T, Alekseyenko A and Wallace K. **A deep learning segmentation strategy that minimizes the amount of manually annotated images [version 2; peer review: 2 approved]** F1000Research 2022, **10**:256 https://doi.org/10.12688/f1000research.52026.2

First published: 30 Mar 2021, **10**:256 https://doi.org/10.12688/f1000research.52026.1

REVISED **Amendments from Version 1**

In this new version, we have changed the first section by:

1. Adding a comparison with a Stardist model trained on the 2018 Data Science Bowl, which is available on Fiji,
2. Better explaining the way we trained U-Net and Mask R-CNN to obtain the results shown in Figure 1.

We also have toned down the benefit of using a conditional GAN to expand the size of the training dataset as it only improves marginally the segmentation accuracy.

Finally, we have completely rewritten the discussion to present observations made in the manuscript rather than a universal guideline. Mainly:

1. The use of publicly available datasets and massive data augmentation are beneficial to build a training dataset and are now common practices in the field,
2. The conditional GAN approach does not improve drastically the segmentation accuracy,
3. Combining instance and semantic segmentations lead to a substantial increase in segmentation accuracy and has the potential to be widely adopted in the field.

Any further responses from the reviewers can be found at the end of the article

Introduction

Over the last decade, deep learning approaches have outperformed all existing methods for image segmentation[1–4]. Semantic segmentation, the estimation of a label at each pixel, and instance segmentation, the identification of individual objects, were successfully applied to spatially characterize biological entities in microscopic images[5–8]. However, these powerful approaches rely on large annotated datasets. While more and more datasets become publicly available[9,10], annotated data for every combination of modalities, tissues and biological objects is far from completion. If the number of images to be segmented is low, a fully manual workflow might be the most time efficient option. Otherwise, procedures to efficiently build training datasets are required to use the full potential of deep learning-based segmentation at a single biological lab scale.

In this paper, we propose a strategy to minimize the amount of time dedicated to manually annotate images and investigate several approaches to maximize accuracy when only using one annotated image. We apply this strategy to segment nuclei stained with DAPI in widefield images of human colorectal adenomas (*i.e.* precancerous polyps) as follows. First, we take advantage of existing training datasets[11,12] and massive data augmentation to obtain a preliminary segmentation. We then use an open source annotation software[12] to manually correct this segmentation and consequently define the training dataset. Next, we simulate synthetic images using a conditional generative adversarial network (GAN)[13] to increase the size of the training dataset. Finally, we combine U-Net[14,15], a semantic segmentation approach, and Mask R-CNN[16], an instance segmentation approach, to improve the nuclear segmentation accuracy.

Methods

Sample preparation

In this study, we used the Medical University of South Carolina (MUSC) pathology laboratory information system CoPath (Cerner Corporation, Kansas City, MO), to identify a convenience sample of colorectal adenomas excised from patients who underwent a sigmoidoscopy or colonoscopy with polypectomy between October 2012 and May 2016. For each patient, we obtained a formalin-fixed, paraffin-embedded (FFPE) tissue block and prepared one H&E and 5, 5-micron sections for immunofluorescence (IF) on FFPE tissue. Prior to the start of the IF procedures, all antibodies were optimized and reviewed by the study immunologist, the pathologist, the epidemiologist, and laboratory personnel to ensure agreement and proper staining. The MUSC Institutional Review Board has approved the research study (IRB # PRO-00007139).

Image acquisition

DAPI was used for nuclear counterstaining. Stained slides were mounted with ProLong™ Gold Antifade Reagent (Cat. # P36934, ThermoFisher) and imaged using the Akoya Vectra® Polaris™ Automated Imaging system (Akoya Biosciences, Marlborough, MA). Whole slide scans were done at 20X magnification and regions of interest where chosen randomly.

Deep learning code

U-Net, Mask R-CNN and pix2pix were coded in Python using the Python libraries numpy[17], tensorflow[18], keras[19], scipy[20] and scikit-image[21].

Training dataset

The training dataset consisted of three 1868 × 1400 images manually annotated with Annotater[12]. Only one image was used to train U-Net and Mask R-CNN as well as pix2pix (conditional GAN) for most of the study, in addition to publicly available datasets (image set BBBC039v1 available from the Broad Bioimage Benchmark Collection[9] and a mouse intestinal epithelium dataset[12]). The two other images were added to the training dataset in the last section to be compared with the combination of results obtained with U-Net and Mask R-CNN (see Figure 3a).

U-Net training

The annotated 1868 × 1400 image was divided into six 622 × 700 images for training: five of these images were included in the training dataset while the last one defined the validation dataset. As U-Net is a semantic segmentation approach, three classes were defined to allow separating nuclei as proposed in 22: inner nuclei, nuclei contours and background. To facilitate nuclei separation, the nuclei contours between touching cells were dilated[22]. To limit over-fitting, the imaging field for images in the training dataset was set to 256 × 256 by randomly cropping the 622 × 700 input images. These cropped images were then normalized to obtain intensity values between 0 and 1. A root mean square prop was used to estimate the parameters of the deep convolutional neural network by minimizing a weighted cross entropy loss to handle class imbalance for 100 epochs without data augmentation and 25 epochs with data augmentation. The weights associated with each class were defined from the training dataset as their inverse proportion. A data augmentation to increase the training dataset by a factor of 100 was processed after normalization with the imgaug python library[23] and included flipping, rotation, pixel dropout, blurring, noise addition and contrast modifications. In Figure 2 and Figure 3, augmented simulated images were obtained by applying the same

modifications with the imgaug python library to simulated images with pix2pix. When combining the annotated image for this study with simulated images and/or existing datasets, the number of augmented images was defined to be balanced between the different data.

U-Net post-processing

An ImageJ macro[24,25] was used to convert the three classes obtained with U-Net to individual nuclei. More specifically, individual nuclei were identified by thresholding the subtraction of the nuclei contours component from the inner nuclei component with a threshold equal to 0.35. A 3D Voronoi tessellation[26] was then applied to assign each pixel to a nucleus. The object component was defined as all pixels whose background component was inferior to 0.95. This object component was then multiplied by the Voronoi tessellation to obtain individual nuclei. The Voronoi tessellation implies that a 1-pixel width area between nuclei is not assigned to any nucleus. To address this problem, the location of these pixels is obtained by subtracting the binary thresholding of the individual nuclei from the object component. The individual nuclei are then dilated[27] and multiplied to this subtraction to be added to the individual nuclei. Finally, nuclei with less than 35 pixels were removed.

Mask R-CNN

The annotated 1868 × 1400 image was divided into thirty-five 266 × 280 images for training: thirty of these images were included in the training dataset while the last five images defined the validation dataset. Version 2.1 of Mask R-CNN[16] was used in this study. The backbone network was defined as the Resnet-101 deep convolutional neural network[28]. We used the code in 5 to define the only class in this study, i.e. the nuclei. A data augmentation to increase the training dataset by a factor of 100 was processed before normalization with the imgaug python library[23] and included resizing, cropping, flipping, rotation, shearing, pixel dropout, blurring, sharpness and brightness modifications, noise addition and contrast modifications. Transfer learning with fine-tuning from a network trained on the coco dataset[29] was also applied. In the first epoch, only the region proposal network, the classifier and mask heads were trained. The whole network was then trained for the next three epochs. In Figure 2 and Figure 3, augmented simulated images were obtained by applying the same modifications with the imgaug python library to simulated images with pix2pix. When combining the annotated image for this study with simulated images and/or existing datasets, the number of augmented images was defined to be balanced between the different data. The maximum image size used for processing Mask R-CNN was larger than 256 as resizing and cropping were applied for data augmentation and set to 512. This parameter was defined as 1024 when other existing datasets were included for training as magnification in these images is higher.

Evaluation

One 1868 × 1400 and one 934 × 1400 manually annotated images were used for evaluation. As proposed in 11, we used the F1 score with respect to the Intersection over Union (*IoU*) to evaluate the different nuclei segmentation approaches. More formally, let $\mathbf{O}_{GT} = \{O_{GT}(e)\}_{e=1,\ldots,n}$ be the set of n ground truth nuclei and $\mathbf{O}_E = \{O_E(e)\}_{e=1,\ldots,m}$ be the set of m estimated nuclei. The *IoU* defined between the truth nucleus $O_{GT}(e_1)$ and the estimated nucleus $O_E(e_2)$ was defined as:

$$IoU(O_{GT}(e_1), O_E(e_2)) = \frac{O_{GT}(e_1) \cap O_E(e_2)}{O_{GT}(e_1) \cup O_E(e_2)}.$$

An $IoU(O_{GT}(e_1), O_E(e_2))$ equal to 0 implies that $O_{GT}(e_1)$ and $O_E(e_2)$ do not share any pixel while an $IoU(O_{GT}(e_1), O_E(e_2))$ equal to 1 means that $O_{GT}(e_1)$ and $O_E(e_2)$ are identical. To ensure that one ground truth nucleus is not associated to multiple estimated nuclei and conversely, we use the following definition for the IoU:

$$IoU^*(O_{GT}(e_1), O_E(e_2)) = \begin{cases} \frac{O_{GT}(e_1) \cap O_E(e_2)}{O_{GT}(e_1) \cup O_E(e_2)} \text{ if } \begin{array}{l} \frac{O_{GT}(e_1) \cap O_E(e_2)}{O_{GT}(e_1) \cup O_E(e_2)} > \frac{O_{GT}(e_1) \cap O_E(e_i)}{O_{GT}(e_1) \cup O_E(e_i)} \forall i \in 1,\ldots,m, \\ \frac{O_{GT}(e_1) \cap O_E(e_2)}{O_{GT}(e_1) \cup O_E(e_2)} > \frac{O_{GT}(e_j) \cap O_E(e_2)}{O_{GT}(e_j) \cup O_E(e_2)} \forall j \in 1,\ldots,n, \end{array} \\ 0 \text{ otherwise.} \end{cases}$$

F1 score for a given IoU^* threshold $t > 0$ can be defined as:

$$F1(t) = \frac{2 \times TP(t)}{2 \times TP(t) + FN(t) + FP(t)},$$

where

$$TP(t) = \sum_{\substack{e_1 \in \{1,\ldots,n\} \\ e_2 \in \{1,\ldots,m\}}} 1(IoU^*(O_{GT}(e_1), O_E(e_2)) > t),$$

$$FN(t) = \sum_{e_1 \in \{1,\ldots,n\}} 1(IoU^*(O_{GT}(e_1), O_E(e_2)) < t), \quad \forall e_2 \in \{1,\ldots,m,\},$$

$$FP(t) = \sum_{e_2 \in \{1,\ldots,m\}} 1(IoU^*(O_{GT}(e_1), O_E(e_2)) < t), \quad \forall e_1 \in \{1,\ldots,n,\},$$

and

$$1(\mathscr{C}) = \begin{cases} 1 \text{ if } \mathscr{C} \text{ is true,} \\ 0 \text{ otherwise.} \end{cases}$$

With a threshold $t = 0.05$, this metric gives the accuracy of a method to identify the correct number of nuclei, while with thresholds in the range 0.05 − 0.9, it evaluates the localization accuracy of the identified nuclear contours.

Conditional GAN

The annotated 1868 × 1400 image was divided into thirty-five 256 × 256 images for training. As defined in 13, U-Net[14] was used for the generator and a convolutional PatchGAN classifier was used for the discriminator. Once trained, nuclei masks had to be generated to simulate images. Distributions for the number of nuclei per image and the size of nuclei were defined from the training dataset. The number of nuclei per image was then modeled as a Gaussian distribution while the size of nuclei was modeled by a Gumbel

distribution to reflect the heavy tail distribution observed in the training dataset. Nuclei masks were then defined as ellipses randomly generated with these distributions with random orientation and a ratio between the two axes defined according to a Gaussian distribution of average s/π and standard deviation of $0.2s/\pi$, where s is the area of the ellipse. 1000 256 × 256 nuclei images were simulated by considering the generated ellipses as nuclei masks.

Combination of instance and semantic segmentations

The combination of results obtained with instance and semantic segmentations was initialized as the nuclei segmented with Mask R-CNN. To prevent from hallucinations, nuclei identified with Mask R-CNN for which the area overlapping with nuclei obtained with U-Net was inferior to 20% were discarded. Then, nuclei identified with U-Net whose area overlapping with nuclei obtained with Mask R-CNN was inferior to 33% were added as new nuclei to the final segmentation. Finally, nuclei with an area inferior to 35 pixels were discarded.

Results

Deep learning-based instance segmentation with existing datasets and massive data augmentation is used to initialize the training dataset

A training dataset is required to train a deep learning method for object segmentation. While new approaches emerge for this task such as interactive machine learning[30], users most often start with manually annotating objects of interest with existing annotation tools[31,32]. As shown in Figure 1a, this task is particularly challenging in our case due to the wide range of morphologies and high density of nuclei in polyps. We use the ImageJ plugin Annotater[12] to efficiently annotate nuclei, a task that takes approximately 30 hours for the image shown in Figure 1a. To avoid a fully manual annotation and save time, it is possible to use the same plugin to correct a nuclei segmentation obtained with an existing method. The watershed method[33], probably the most used method for nuclei segmentation in fluorescence microscopy images, correctly identifies a high number of nuclei (high F1 score for a low IoU threshold in Figure 1). Unfortunately, under- and over-segmentations, a well-known limitation of this approach, lead to a poor segmentation localization (rapidly decreasing F1 score with increasing IoU thresholds in Figure 1). Alternatively, pretrained deep learning models for nuclei segmentation are available. Stardist[7], one of the most popular approaches in microscopy, can be processed as a Fiji plugin[25] with a model trained on the 2018 Data Science Bowl[10]. While the number of nuclei correctly identified is lower than with the watershed method (lower F1 score for low IoU thresholds in Figure 1b–c), their localization accuracy is much higher (higher F1 score for high IoU thresholds in Figure 1b–c). Another possibility is to train deep learning approaches with existing training datasets. We propose to train a U-Net model and a Mask R-CNN model with a high throughput chemical screen on U2OS cells dataset (CC) (image set BBBC039v1 available from the Broad Bioimage Benchmark Collection[9]) and a widefield mouse intestinal epithelium dataset (MIE)[12]. These models are then used to segment the image shown in Figure 1a. While U-Net demonstrates a poor performance (Figure 1b), Mask RCNN clearly surpasses the watershed approach and the pretrained Stardist model (Figure 1c). When compared to the latter, the good performance of Mask R-CNN is explained by the fact that the MIE dataset includes epithelial nuclei, even though they come from mice. Correcting this segmentation with Annotater takes about 15–20 hours, which is clearly faster than an annotation from scratch. Training the Mask R-CNN model with the CC and MIE datasets takes about 12 hours but has the great advantage that it does not require human interaction. For both UNet and Mask R-CNN, a massive data augmentation (100 times) clearly improves the performance.

Increasing the training dataset by using a conditional GAN improves nuclear segmentation accuracy

When only considering the annotated image in Figure 1a in the training dataset, U-Net leads to higher segmentation accuracy than Mask R-CNN (Figure 2a-b). To increase the training dataset, we use the same annotated image to train a conditional Generative Adversarial Network (GAN)[13] and simulate images showing nuclei from masks defined as random ellipses generated with the distributions of nuclei size and nuclei number observed in the training dataset (see Figure 2c and Methods). Only using simulated images lead to a lower accuracy for both deep learning approaches, even though applying mathematical operations to these synthetic images (augmented simulated training dataset, see Methods) improves the segmentation accuracy. However, pooling together augmented simulated images and the annotated image from Figure 1a slightly improves U-Net performance and distinctly increases the number of accurately identified nuclei with Mask R-CNN while decreasing the segmentation localization precision. Finally, adding existing datasets clearly leads to the optimal results for Mask R-CNN while degrading the accuracy for U-Net, which is consistent with the inability for this approach to generalize nuclear segmentation for different data as shown in Figure 1b. Overall, U-Net marginally benefits from using simulated images (red curve versus black curve in Figure 1a) while the main gain for Mask R-CNN comes from the use of CC/MIE datasets and data augmentation (orange curve versus red and black curves in Figure 1b).

Combining semantic and instance segmentations improves nuclear segmentation accuracy

Nuclei segmented with Mask R-CNN show a higher localization precision than those obtained with U-Net as shown in Figure 1a-b. However, nuclei that are harder to delineate are missed with Mask R-CNN while U-Net accurately identifies pixels that belong to nuclei, even though the separation between individual nuclei might not be precise. In order to get the best of both worlds, we propose to combine the results obtained with U-Net trained with one annotated image with data augmentation and augmented simulated images, and the results obtained with Mask R-CNN trained with one annotated image with data augmentation, augmented simulated images and existing datasets with data augmentation (see Methods). As shown in Figure 3, these results demonstrate a higher F1 score for any IoU threshold than obtained with U-Net or Mask R-CNN trained with 3 times more annotated images. The corresponding segmented nuclei are shown in Figure 4.

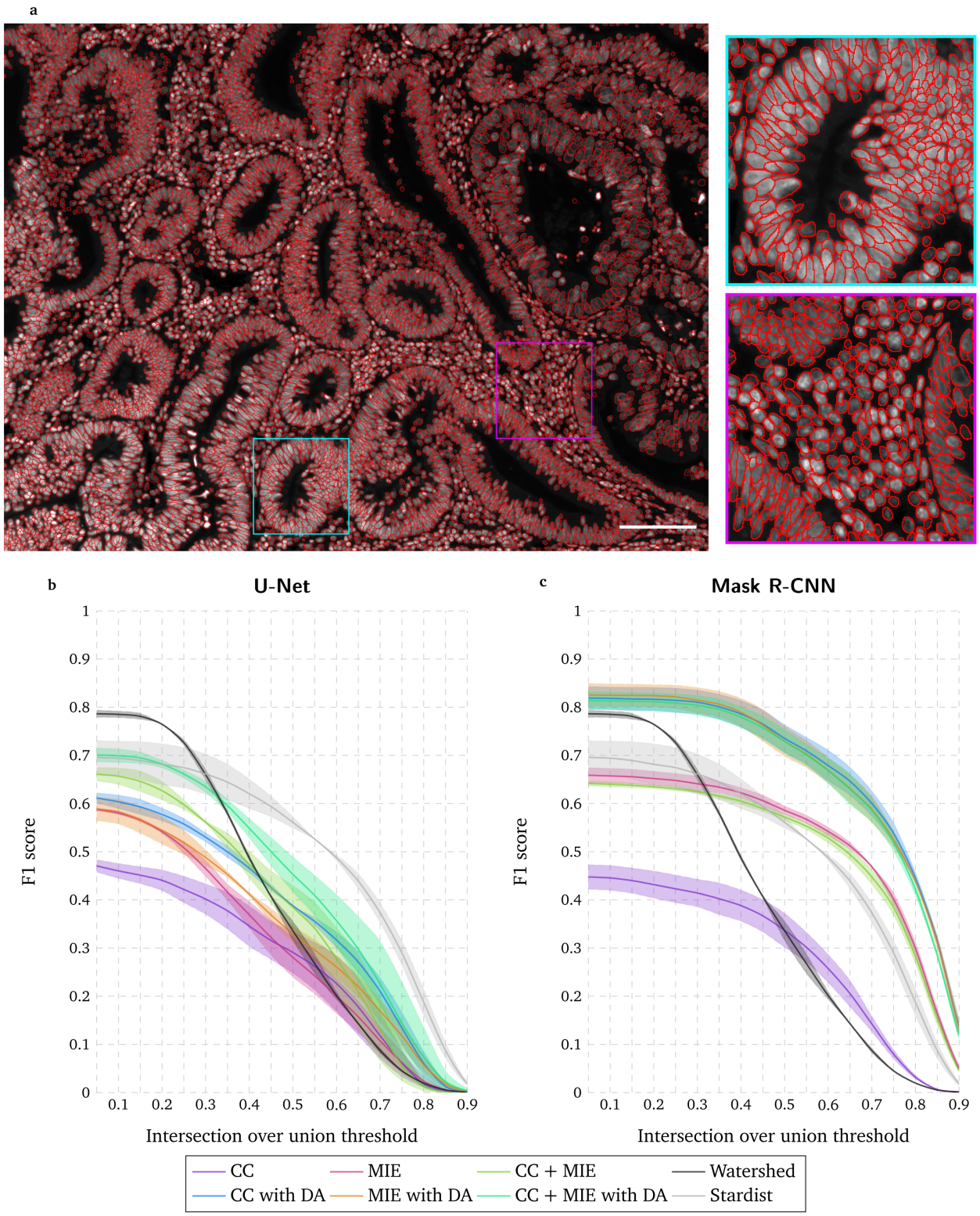

Figure 1. Manual annotation and evaluation of deep learning-based segmentation with existing training datasets. a Widefield acquisition of a human polyp biopsy stained with DAPI. Manually annotated nuclei are overlaid as red circles. Zoomed-in regions are displayed on the right side with corresponding squared colors. Scale bar = 100μm. **b**–**c** F1 score for range of IoU thresholds obtained with the watershed method, with Stardist, with U-Net **b** and Mask R-CNN **c** approaches trained with a high-throughput chemical screen on U2OS cells dataset (CC) or/and a widefield mouse intestinal epithelium dataset (MIE), with and without data augmentation (DA). Lines correspond to average F1 score over the two tested images while the shaded areas represent the standard deviation.

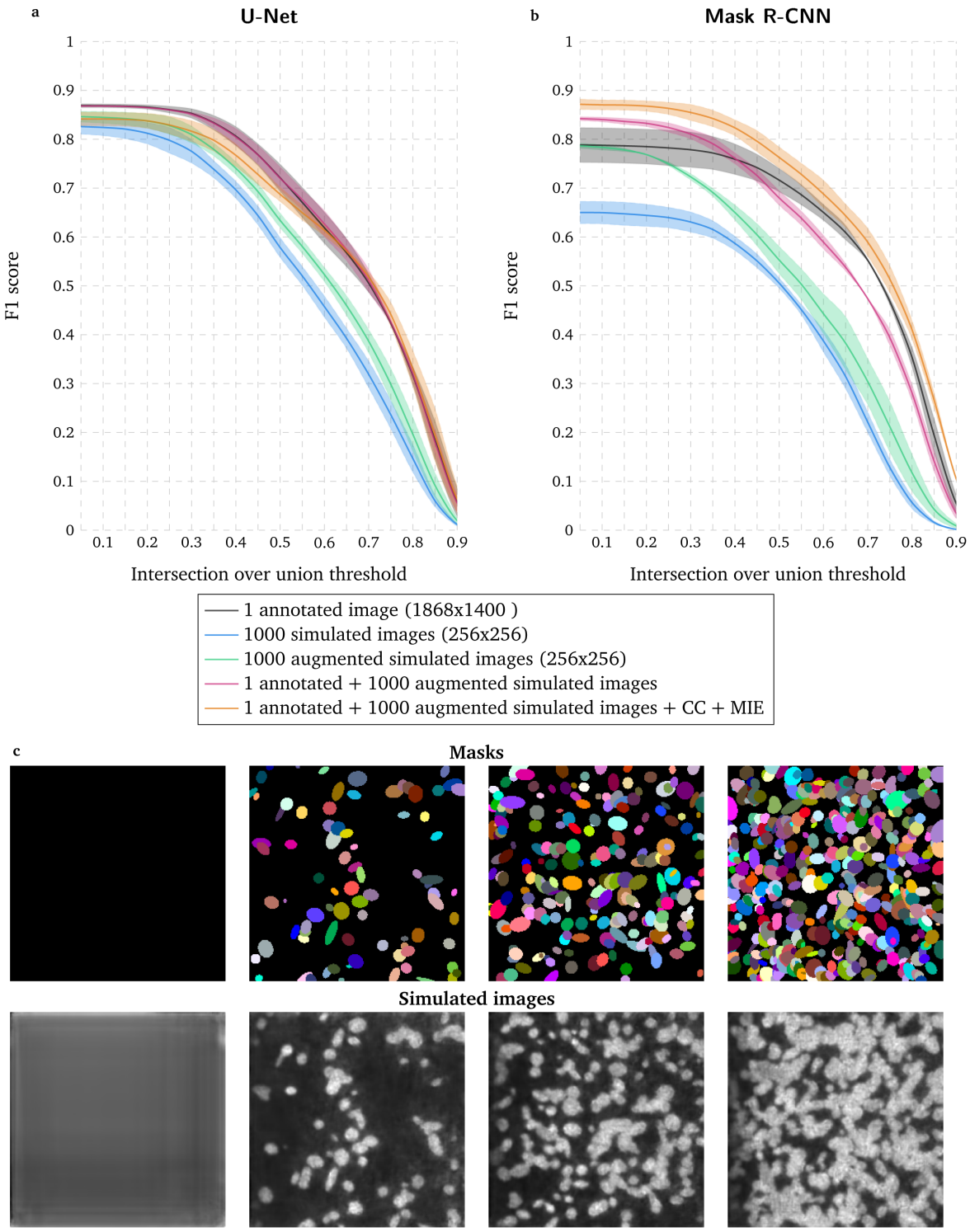

Figure 2. Evaluation of deep learning-based segmentation when using a conditional Generative Adversarial Network to increase the size of the training dataset. a First row: masks generated as ellipses (see Methods) and represented with unique colors. Second row: images simulated from masks shown in first row with a conditional Generative Adversarial Network (GAN). **b**–**c** F1 score for range of IoU thresholds obtained with U-Net **b** and Mask R-CNN **c** trained with 1 annotated image with data augmentation (DA), 1000 simulated images, 1000 augmented simulated images, 1 annotated image with DA combined with 1000 augmented simulated images and 1 annotated image with DA combined with 1000 augmented simulated images as well as a high-throughput chemical screen on U2OS cells dataset (CC) and a widefield mouse intestinal epithelium dataset (MIE). Lines correspond to average F1 score over the two tested images while the shaded areas represent the standard deviation.

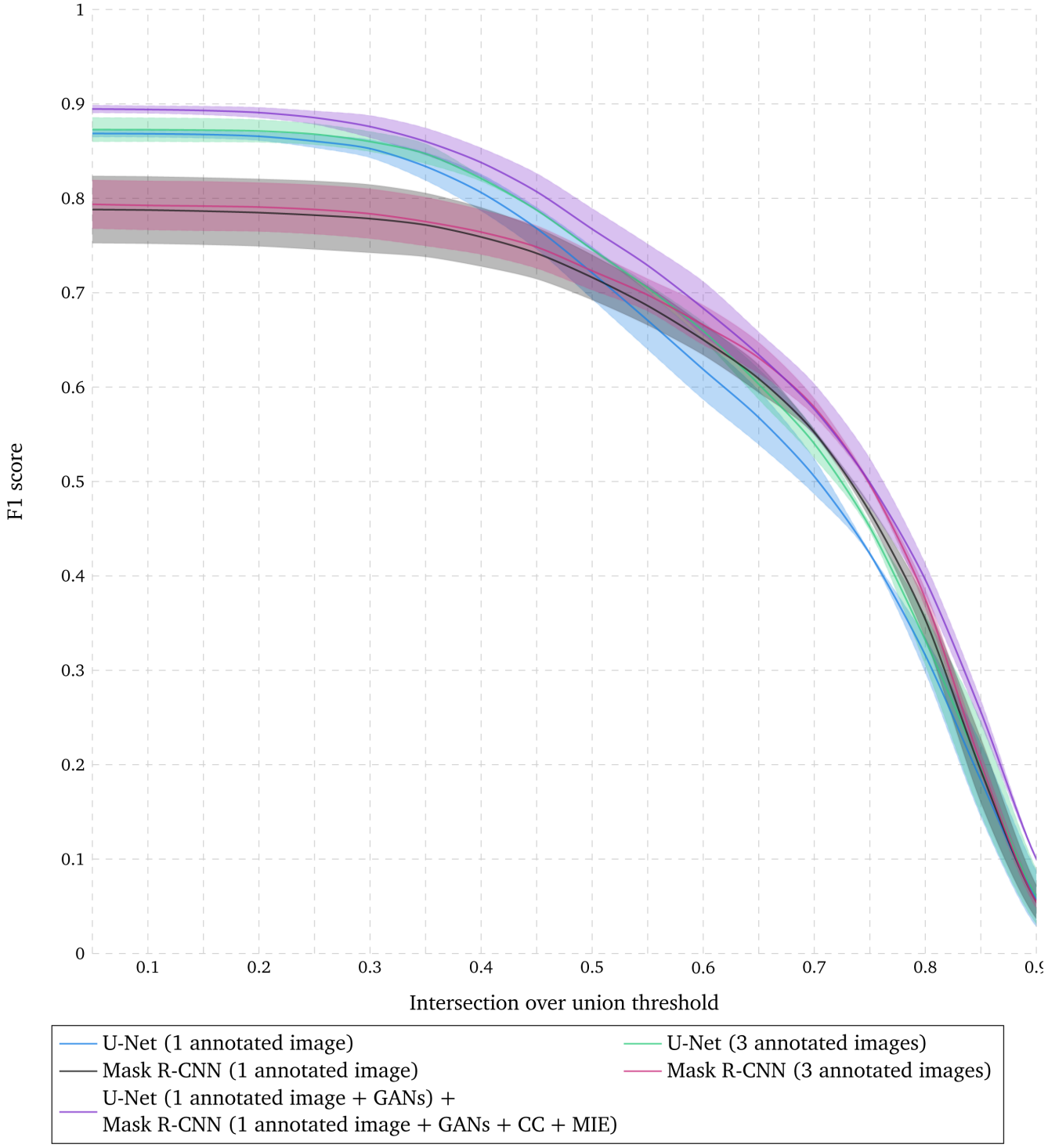

Figure 3. Evaluation of nuclear segmentation when combining U-Net and Mask R-CNN. F1 score for range of IoU thresholds obtained with U-Net trained with 1 and 3 annotated images with data augmentation (DA), Mask R-CNN trained with 1 and 3 annotated images with DA, and the combination of results obtained with U-Net trained with 1 annotated image with DA and augmented simulated images, and the results obtained with Mask R-CNN trained with 1 annotated image with DA, augmented simulated images and existing datasets with DA. Lines correspond to average F1 score over the two tested images while the shaded areas represent the standard deviation.

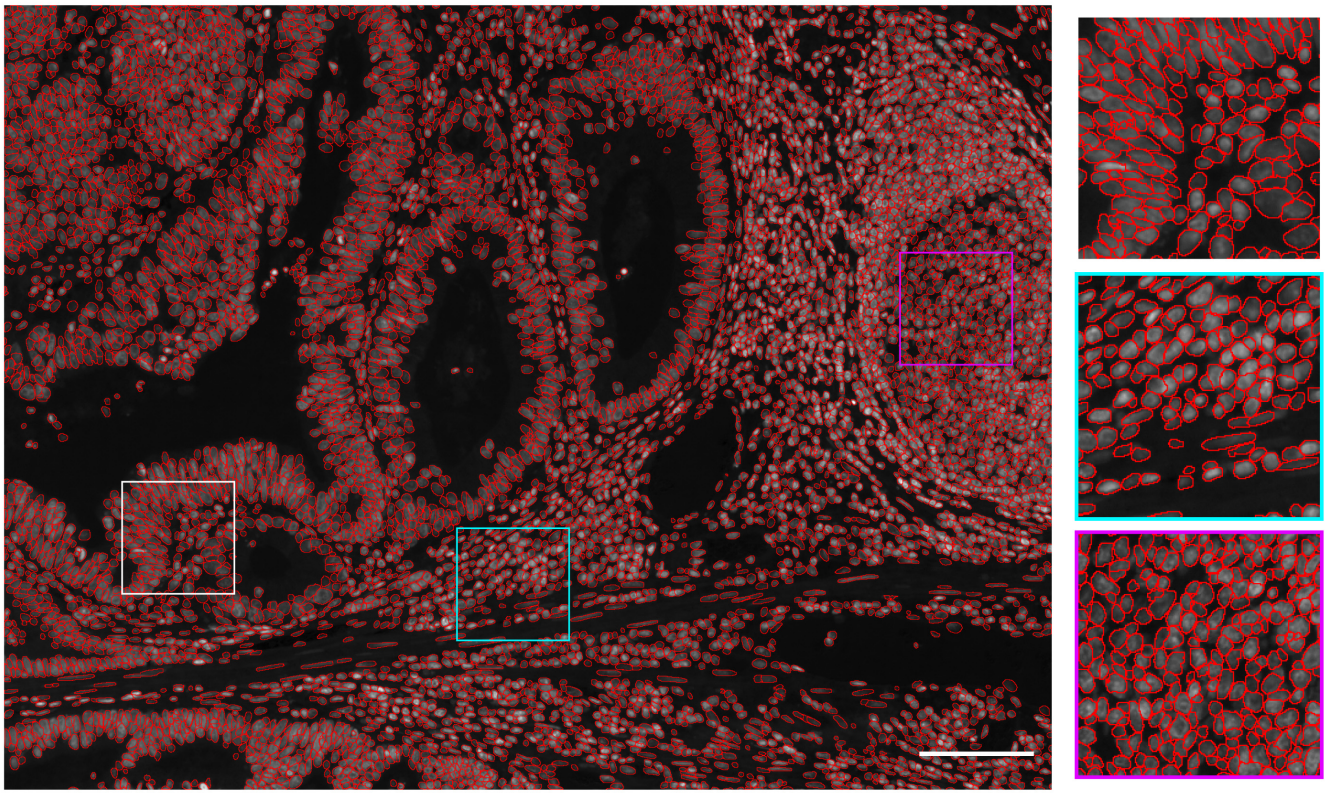

Figure 4. Nuclear segmentation example when combining U-Net and Mask R-CNN. Segmented nuclei obtained by combining U-Net and Mask R-CNN overlaid as red circles over the processed image. Zoomed-in regions are displayed on the right side with corresponding squared colors. Scale bar = 100μm.

Discussion

This study explores several strategies to minimize the amount of manually annotated data required to successfully train a deep learning model for instance segmentation. As already established in the field, the use of existing training datasets, even though modalities and/or tissues differ, allows to train instance segmentation models and obtain results on targeted data that can be manually corrected to initialize a new training dataset. The use of massive data augmentation is another well known approach to drastically increase the segmentation accuracy. While the use of conditional GANs to expand the size of the training dataset seems promising, the gain in accuracy shown in this study is not very convincing. The simulation pipeline used to generate the masks might have been too simplistic. In particular, the variety of nuclei shapes could be enriched. Finally, combining semantic and instance segmentation results leads to a substantial increase in segmentation accuracy. While unusual in the field, we believe that this method has the potential to become more common in the community. Combining these strategies enables to remarkably reduce the amount of data to be manually annotated, waiting for methods that offer the promise to eliminate this time consuming task, such as self- and partially supervised methods, currently in development.

Data availability

The five annotated images are available at https://github.com/tpecot/DeepLearningBasedSegmentationForBiologists/tree/main/Data/AnnotatedNuclei. This project contains the following data:

- *Polyp12_[10837,39273]_component_data.tiff*: image used for training U-Net and Mask R-CNN in all figures and for training pix2pix in Figure 2
- *Polyp40_[13694,34105]_component_data.tiff and Polyp42_[12011,37598]_component_data.tiff*: two images used for training U-Net and Mask R-CNN in Figure 3
- *Polyp12_[12699,39273]_component_data.tiff and Polyp42_[12942,36900]_component_data.tiff*: two images used for evaluation in all figures

The images generated with pix2pix and used for training U-Net and Mask R-CNN in Figure 2–Figure 3 are available at https://github.com/tpecot/NucleiSimulationWithConditionalGAN/tree/main/datasets/Nuclei_polyps_1image.

Software availability

The code with the parameters used to train and process all experiments presented in this manuscript with U-Net and

Mask R-CNN is available at https://github.com/tpecot/DeepLearningBasedSegmentationForBiologists/tree/main/Codes.

Archived code as at time of publication: https://doi.org/10.5281/zenodo.4608795[34]

License: GPL3

The code with the parameters used to train and generate images with pix2pix is available at https://github.com/tpecot/NucleiSimulationWithConditionalGAN.

Archived code as at time of publication: https://doi.org/10.5281/zenodo.4608793[35]

License: GPL3

The ImageJ macro used to convert the output classes obtained with U-Net to individual nuclei is available at https://github.com/tpecot/DeepLearningBasedSegmentationForBiologists/tree/main/Codes.

Archived macro as at time of publication: https://doi.org/10.5281/zenodo.4608795[34]

License: GPL3

Acknowledgements

This publication was supported by COST Action NEU-BIAS (CA15124), funded by COST (European Cooperation in Science and Technology).

References

1. Krizhevsky A, Sutskever I, Hinton GE: **Imagenet classification with deep convolutional neural networks.** *Adv Neural Inf Process Syst.* 2012; 1097–1105.
Reference Source
2. Cireşan D, Meier U, Schmidhuber J: **Multi-column deep neural networks for image classification.** In *2012 IEEE conference on computer vision and pattern recognition.* IEEE, 2012; 3642–3649.
Reference Source
3. LeCun Y, Bengio Y, Hinton G: **Deep learning.** *Nature.* 2015; **521**(7553): 436–444.
PubMed Abstract | Publisher Full Text
4. Schmidhuber J: **Deep learning in neural networks: An overview.** *Neural Netw.* 2015; **61**: 85–117.
PubMed Abstract | Publisher Full Text
5. Hollandi R, Szkalisity A, Toth T, *et al.*: **nucleaizer: a parameter-free deep learning framework for nucleus segmentation using image style transfer.** *Cell Systems.* 2020; **10**(5): 453–458.e6.
Publisher Full Text
6. Moen E, Bannon D, Kudo T, *et al.*: **Deep learning for cellular image analysis.** *Nat Methods.* 2019; **16**(12): 1233–1246.
PubMed Abstract | Publisher Full Text
7. Schmidt U, Weigert M, Broaddus C, *et al.*: **Cell detection with star-convex polygons.** In *International Conference on Medical Image Computing and Computer-Assisted Intervention.* Springer, 2018; 265–273.
Publisher Full Text
8. Mandal S, Uhlmann V: **Splinedist: Automated cell segmentation with spline curves.** *bioRxiv.* 2020.
Publisher Full Text
9. Ljosa V, Sokolnicki KL, Carpenter AE: **Annotated high-throughput microscopy image sets for validation.** *Nat Methods.* 2012; **9**(7): 637.
PubMed Abstract | Publisher Full Text | Free Full Text
10. Caicedo JC, Goodman A, Karhohs KW, *et al.*: **Nucleus segmentation across imaging experiments: the 2018 data science bowl.** *Nat Methods.* 2019; **16**(12): 1247–1253.
PubMed Abstract | Publisher Full Text | Free Full Text
11. Caicedo JC, Roth J, Goodman A, *et al.*: **Evaluation of deep learning strategies for nucleus segmentation in fluorescence images.** *Cytometry A.* 2019; **95**(9): 952–965.
PubMed Abstract | Publisher Full Text | Free Full Text
12. Pécot T, Cuitiño MC, Johnson RH, *et al.*: **Deep learning tools and modeling to estimate the temporal expression of cell cycle from 2D still images.** *bioRxiv.* 2021.
Publisher Full Text
13. Isola P, Zhu JY, Zhou T, *et al.*: **Image-to-image translation with conditional adversarial networks.** In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017; 1125–1134.
Publisher Full Text
14. Ronneberger O, Fischer P, Brox T: **U-Net: Convolutional Networks for Biomedical Image Segmentation.** In *International Conference on Medical image computing and computer-assisted intervention.* Springer, 2015; 234–241.
Publisher Full Text
15. Falk T, Mai D, Bensch R, *et al.*: **U-net: deep learning for cell counting, detection, and morphometry.** *Nat Methods.* 2019; **16**(1): 67–70.
PubMed Abstract | Publisher Full Text
16. He K, Gkioxari G, Dollár P, *et al.*: **Mask r-cnn.** In *Proceedings of the IEEE international conference on computer vision*. 2017; 2961–2969.
Reference Source
17. van der Walt S, Colbert SC, Varoquaux G: **The numpy array: a structure for efficient numerical computation.** *Comput Sci Eng.* 2011; **13**(2): 22–30.
Publisher Full Text
18. Abadi M, Agarwal A, Barham P, *et al.*: **TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems**. 2015.
Reference Source
19. Chollet F: **Keras**. 2015.
Reference Source
20. Virtanen P, Gommers R, Oliphant TE, *et al.*: **SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python.** *Nat Methods.* 2020: **17**(3): 261–272.
PubMed Abstract | Publisher Full Text | Free Full Text
21. van der Walt S, Schönberger JL, Nunez-Iglesias J, *et al.*: **scikit-image: image processing in python.** *PeerJ.* 2014; **2**: e453.
PubMed Abstract | Publisher Full Text | Free Full Text
22. Van Valen DA, Kudo T, Lane KM, *et al.*: **Deep learning automates the quantitative analysis of individual cells in live-cell imaging experiments.** *PLoS Comput Biol.* 2016; **12**(11): e1005177.
PubMed Abstract | Publisher Full Text | Free Full Text
23. Jung AB, Wada K, Crall J, *et al.*: **imgaug**. 2020; accessed 01-Feb-2020.
Reference Source
24. Schneider CA, Rasband WS, Eliceiri KW: **Nih image to imagej: 25 years of image analysis.** *Nat Methods.* 2012; **9**(7): 671–675.
PubMed Abstract | Publisher Full Text | Free Full Text
25. Schindelin J, Arganda-Carreras I, Frise E, *et al.*: **Fiji: an open-source platform for biological-image analysis.** *Nat Methods.* 2012; **9**(7): 676–682.
PubMed Abstract | Publisher Full Text | Free Full Text
26. Ollion J, Cochennec J, Loll F, *et al.*: **Tango: a generic tool for high-throughput 3d image analysis for studying nuclear organization.** *Bioinformatics.* 2013; **29**(14): 1840–1841.
PubMed Abstract | Publisher Full Text | Free Full Text
27. Legland D, Arganda-Carreras I, Andrey P: **Morpholibj: integrated library and plugins for mathematical morphology with imagej.** *Bioinformatics.* 2016; **32**(22): 3532–3534.
PubMed Abstract | Publisher Full Text
28. He K, Zhang X, Ren S, *et al.*: **Deep residual learning for image recognition.** In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016; 770–778.
Publisher Full Text
29. Lin TY, Maire M, Belongie S, *et al.*: **Microsoft coco: Common objects in context.** In *European conference on computer vision.* Springer, 2014; **8693**: 740–755.
Publisher Full Text
30. Ouyang W, Le T, Xu H, *et al.*: **Interactive biomedical segmentation tool**

powered by deep learning and imjoy. [version 1; peer review: 1 approved, 1 approved with reservations]. *F1000Res.* 2021; **10**(142): 142.
Publisher Full Text

31. Bankhead P, Loughrey MB, Fernández JA, *et al.*: **Qupath: Open source software for digital pathology image analysis.** *Sci Rep.* 2017; **7**(1): 16878.
PubMed Abstract | Publisher Full Text | Free Full Text

32. Sofroniew N, Lambert T, Evans K, *et al.*: **napari**. 2021.
Publisher Full Text

33. Vincent L, Soille P: **Watersheds in digital spaces: an efficient algorithm based on immersion simulations.** *IEEE Trans Pattern Anal Mach Intell.* 1991; **13**(6): 583–598.
Publisher Full Text

34. Pécot T: **Deep Learning-based segmentation for biologists.** 2021.
http://www.doi.org/10.5281/zenodo.4608793

35. Pécot T: **Nuclei Simulation with Conditional GAN.** 2021.
http://www.doi.org/10.5281/zenodo.4608795

Open Peer Review

Current Peer Review Status: ✔ ✔

Version 2

Reviewer Report 24 January 2022

https://doi.org/10.5256/f1000research.115746.r120021

© **2022 Laine R.** This is an open access peer review report distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

✔ **Romain F Laine**
MRC Laboratory for Molecular Cell Biology, University College London, London, WC1E 6BT, UK

The authors have now fully addressed my comments. This manuscript makes a nice and timely story, I have enjoyed reading and helping reviewing it. Nice work!

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: I am a quantitative imaging specialist, focused on fluorescence microscopy, super-resolution, and quantitative analysis method development, including deep learning.

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Reviewer Report 24 January 2022

https://doi.org/10.5256/f1000research.115746.r120022

© **2022 Lucas A.** This is an open access peer review report distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

✔ **Alice Lucas**
Broad Institute, Cambridge, MA, USA

Thank you for addressing my earlier comments.

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Deep Learning, Computer Vision, Image and Video Processing

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Version 1

Reviewer Report 14 December 2021

https://doi.org/10.5256/f1000research.55252.r101244

© 2021 Laine R. This is an open access peer review report distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

?

Romain F Laine
MRC Laboratory for Molecular Cell Biology, University College London, London, WC1E 6BT, UK

Pecot *et al* present a nice set of ideas about how to improve the pipeline of nuclei segmentation. The premise of this work is that it is time consuming to generate good quality annotation for DL training. The authors are absolutely right here, it takes time and can be discouraging.

The authors test a couple of interesting approaches to help with that:

1. Use of large openly accessible dataset to create pretrained dataset, as is commonly done in the field.

2. Use of data augmentation to improve generalization of the model, also commonly done in the field already.

3. The use of a generator model (here pix2pix as a conditional GAN), to expand the size f the training dataset.

4. Combine output of 2 common segmentation networks (U-Net and MaskRCNN) to improve accuracy.

The points 1 and 2 are already well established in the field and will be systematically done nowadays, with almost any DL networks when data is available. Segmentation dataset are available as the authors highlight. So these aspects are sanity checks here and not novel implementations. However, it is reassuring to see that augmentation and use of pretrained models are helpful here as well.

The more interesting aspects of this work lie in the use of GAN for expanding the size of the training dataset from an annotated image and the combination of output. Although the use of GAN makes sense for this application, the gains from such approach are clearly quite marginal as can be seen on Figure 2a and 2b comparing the black and red lines, while the main gains are again from the use of additional dataset and augmentation as observed on Figure 1. It's an important observation but maybe not as essential to the pipeline as is described in the manuscript as it stands. I suggest toning down the importance of this and clearly highlighting that the gains

are in fact low here.
Maybe the authors could further discuss why the gains are only small here: maybe the simulation pipeline from the masks to generate the training dataset of pix2pix is too simplistic for instance, wider range of shapes, background lights, heterogeneity of intensities or patterns on the nuclei etc.

On the contrary it's quite clear that the combination of U-Net and MaskRCNN output are beneficial to the overall performance of the method and that's nicely shown here. I think that combining DL model outputs is currently underused and this is a nice additional demonstration of this here.

As additional comments, I would highlight a couple of additional work that are missing from the context described in this paper:

- Kaibu (Wei Ouyang *et al* F1000, https://f1000research.com/articles/10-142) is an interactive tool for simultaneous training of segmentation models and segmentation, this circumvents a range of issues mentioned here, it should be mentioned.

- StarDist (https://github.com/stardist/stardist) from Uwe Schmidt and Martin Weigert, is an excellent tool for nuclei segmentation and is not included here. I suggest that the authors compare their IoU curves to those obtained from the pre-trained models provided by the method (even as Fiji plugin). This will give the readers a baseline on which to compare the approaches described here, which still require an investment in time to train multiple models and annotations

- The cost/benefit analysis of manual annotation vs automated (DL based or not) should be mentioned, it's not always worth doing DL for that, it often depends on the size of the dataset to be segmented.

- Although having an annotator GUI and some tools to get some improvement on segmentation performance are important today, a large efforts is now put into approaches that are self-supervised or partially supervised, which would circumvent the issues of annotation time altogether. These are not currently available to the wide bioimaging community but should be mentioned in conclusion, looking at the future of segmentation pipelines.

Overall, I think that it is a nice piece of work describing the performance of a range of approaches in a systematic and clear manner, which are useful to the bioimaging community. However, they are presented as guidelines to building a segmentation pipeline and I would not think that as such, it describes the general thoughts about the matter in the community. I'd consider rewording the conclusions focusing on the observations of the tests the authors made rather than presenting it as a universal guideline.

References

1. Ouyang W, Le T, Xu H, Lundberg E: Interactive biomedical segmentation tool powered by deep learning and ImJoy. *F1000Research*. 2021; **10**. Publisher Full Text

Is the rationale for developing the new method (or application) clearly explained?

Yes

Is the description of the method technically sound?

Yes

Are sufficient details provided to allow replication of the method development and its use by others?
Yes

If any results are presented, are all the source data underlying the results available to ensure full reproducibility?
Yes

Are the conclusions about the method and its performance adequately supported by the findings presented in the article?
Partly

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: I am a quantitative imaging specialist, focused on fluorescence microscopy, super-resolution, and quantitative analysis method development, including deep learning.

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Author Response 17 Dec 2021

Thierry Pecot, Medical University of South Carolina, Charleston, USA

We thank Romain Laine for his enlightening observations.
To answer Dr Laine remarks, we changed the manuscript accordingly:

- We added a sentence about the cost/benefit analysis of manual annotation vs automated in the introduction.
- We mentioned the use of interactive machine learning and cited Kaibu (Wei Ouyang *et al* F1000, https://f1000research.com/articles/10-142) at the beginning of the first section.
- We added a comparison to Stardist trained with the 2018 Science Bowl (Fiji plugin) in Fig.1. We then compared its performance to the watershed approach, to U-Net and to Mask R-CNN trained with the CC/MIE datasets in the first section.
- We completely changed the discussion, focusing on the observations made in the manuscript. More particularly, we acknowledge that the use of publicly available datasets and massive data augmentation are beneficial to build a training dataset and are now common practices in the field. We also underline the disappointing accuracy obtained when using pix2pix (we also changed the end of section 2 accordingly). We emphasize the interest of combining instance and semantic segmentations. We finally introduce self- and partially supervised methods that offer

the promise to eliminate manual annotation.

Competing Interests: No competing interests were disclosed.

Reviewer Report 03 August 2021

https://doi.org/10.5256/f1000research.55252.r89422

© 2021 Lucas A. This is an open access peer review report distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Alice Lucas
Broad Institute, Cambridge, MA, USA

The authors propose multiple strategies to improve segmentation results given a new dataset.

Instead of manually annotating a training image from scratch, the authors recommend to leverage knowledge learned by networks pre-trained on other larger datasets. Therefore, they propose to first train a model on large existing datasets (that differ from the final dataset of interest). The trained model is then used to annotate the desired training image, and these predictions are then manually corrected using Annotater. This allows the authors to manually annotate for 15-20 hours, compared with 30 hours when annotating the image from scratch.

A second solution that they implement in order to improve their final segmentation results is to (1) train a conditional GAN on their annotated image and (2) use the cGAN to predict additional synthesized segmentation masks. The UNet and Mask-RCNN can then use this additional data for training.

Finally, to further improve their results, they combine results obtained from their trained UNet and their trained Mask-RCNN to obtain a final instance segmentation map. More specifically, the semantic segmentations from UNet are post-processed to obtain instance segmentation masks, and merged (following a specific protocol) with those predicted by the trained Mask-RCNN.

A few comments:

- Clarity regarding the purposes of the different training sets used could be improved. At first it was not clear to me how the CC/MIE datasets related to the final training dataset of interest (the 1868 x 1400 image). It could be made a bit more explicit that (1) the CC / MIE data is used to pre-train a neural network, (2) this neural network is then applied on the image of interest to provide the annotations, and (3) final training data is obtained by correcting these predictions. (4) On this final training data will be trained UNet and Mask-RCNN.

- The text “Only one image was used to train [...]” is a bit of a misleading statement. In the

end, when looking at the whole pipeline, a very large dataset of annotated images was used to get to these results.

- It would be interesting to know how many hours it took to pre-train Mask-RCNN and UNet on the large datasets, as well as for training the conditional GAN. This is helpful especially for better comparing the 30 hours of manual annotation from scratch vs. the 15-20 hours when using these strategies.

Is the rationale for developing the new method (or application) clearly explained?
Yes

Is the description of the method technically sound?
Yes

Are sufficient details provided to allow replication of the method development and its use by others?
Yes

If any results are presented, are all the source data underlying the results available to ensure full reproducibility?
Yes

Are the conclusions about the method and its performance adequately supported by the findings presented in the article?
Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Deep Learning, Computer Vision, Image and Video Processing

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Author Response 17 Dec 2021

Thierry Pecot, Medical University of South Carolina, Charleston, USA

We thank Alice Lucas for her insightful remarks and apologize for the delay between her review and our response, we were waiting for a second reviewer.

To answer Dr Lucas comments, we changed the manuscript accordingly:

- We rephrased the first section to better explain what was done. More specifically, U-Net and Mask R-CNN are trained with CC/MIE datasets along with a massive data augmentation. The trained models are then used to segment the image shown in Fig.1 **a**. The accuracy obtained with these models is compared to the watershed approach and to a Stardist model trained with the 2018 Data Science Bowl. As Mask

R-CNN demonstrates the most accurate results, the segmented nuclei obtained with this approach are then manually corrected to initialize a training dataset.

- To clarify the misleading text "Only one image was used to train [...]" , we changed the **Training dataset** section in **Methods** and added that publicly available datasets were used in addition to the manually annotated image of human precancerous polyp biopsy.

- We added a sentence about the time taken to train a Mask R-CNN model on the CC/MIE datasets at the end of the first section to better compare the 30 hours of manual annotation from scratch vs. the 15-20 hours when using this strategy.

Competing Interests: No competing interests were disclosed.

The benefits of publishing with F1000Research:

- Your article is published within days, with no editorial bias
- You can publish traditional articles, null/negative results, case reports, data notes and more
- The peer review process is transparent and collaborative
- Your article is indexed in PubMed after passing peer review
- Dedicated customer support at every stage

For pre-submission enquiries, contact research@f1000.com

F1000Research