

# Revisiting foundation models for cell instance segmentation

Anwai Archit<sup>1</sup> 

ANWAI.ARCHIT@UNI-GOETTINGEN.DE

Constantin Pape<sup>1,2,3</sup> 

CONSTANTIN.PAPE@INFORMATIK.UNI-GOETTINGEN.DE

<sup>1</sup> *Georg-August-University Göttingen, Institute of Computer Science*

<sup>2</sup> *CAIMed - Lower Saxony Center for AI & Causal Methods in Medicine, Göttingen*

<sup>3</sup> *Cluster of Excellence Multiscale Bioimaging (MBExC), Georg-August-University Göttingen*

**Editors:** Under Review for MIDL 2026

## Abstract

Cell segmentation is a fundamental task in microscopy image analysis. Several foundation models for cell segmentation have been introduced, virtually all of them are extensions of Segment Anything Model (SAM), improving it for microscopy data. Recently, SAM2 and SAM3 have been published, further improving and extending the capabilities of general-purpose segmentation foundation models. Here, we comprehensively evaluate foundation models for cell segmentation (CellPose-SAM, CellSAM,  $\mu$ SAM) and for general-purpose segmentation (SAM, SAM2, SAM3) on a diverse set of (light) microscopy datasets, for tasks including cell, nucleus and organoid segmentation. Furthermore, we introduce a new segmentation strategy called automatic prompt generation (APG) that can be used to further improve the segmentation quality of SAM-based microscopy foundation models. APG achieves state-of-the-art results on several of the benchmarked datasets, providing practical benefits for microscopy image analysis. Moreover, our work provides important lessons for adaptation strategies of SAM-style models to microscopy and provides a strategy for creating even more powerful microscopy foundation models.

**Keywords:** foundation models, cell segmentation, segment anything model

## 1. Introduction

Instance segmentation is one of the most important image analysis tasks in microscopy, enabling phenotypic drug screening in high-content imaging (Chandrasekaran et al., 2024), analysis of embryogenesis at the cellular level (Lange et al., 2024), and many other applications. Virtually all current methods for microscopy instance segmentation are based on deep learning, such as dedicated tools addressing cell (Stringer et al., 2021) and nucleus (Schmidt et al., 2018) segmentation in light microscopy, nucleus segmentation in histopathology (Graham et al., 2019), or segmentation of mitochondria (Conrad and Narayan, 2023) and other organelles (Muth et al., 2025) in electron microscopy. A repository (Ouyang et al., 2022) collects trained models for such segmentation tasks, compatible with tools for model inference (Gómez-de Mariscal et al., 2021; Berg et al., 2019).

These dedicated models can accelerate many analyses, yet the large number of tools and the potential lack of pretrained models for specific tasks put a burden on users without substantial computational expertise. Hence, foundation models have been introduced in this domain, enabling a wider range of tasks with a single model (Archit et al., 2025a; Pachitariu et al., 2025; Israel et al., 2023; Hörst et al., 2024; Griebel et al., 2025). They

are predominantly based on the Segment Anything Model (SAM) (Kirillov et al., 2023), a general-purpose segmentation foundation model. SAM itself has been extended to video data by SAM2 (Ravi et al., 2024) and, recently, to text- and example-based segmentation by SAM3 (Carion et al., 2025). The latter also included microscopy data in its training set.

These developments open up the following questions: (i) What is the best strategy for adapting a SAM-style model to microscopy? (ii) Are specific (foundation) models for microscopy segmentation still needed or do general-purpose segmentation models, in particular SAM3, make them obsolete? (iii) What influence does the training data (modalities, size, data diversity) have on model performance?

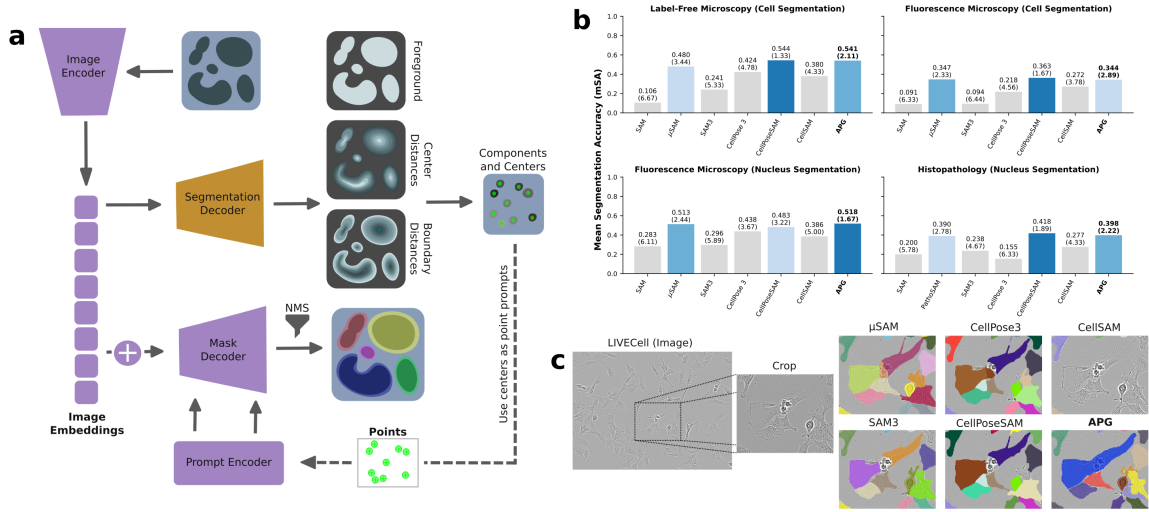


Figure 1: a) Overview of our new instance segmentation method, automatic prompt generation (APG), which re-purposes the trained  $\mu$ SAM (and PathoSAM) models by deriving point prompts from decoder predictions, predicting masks based on these prompts, and then filtering overlapping masks via NMS. b) Overview of segmentation results for four different microscopy modalities. We report the averaged rank over the 9 datasets per domain in parantheses, top three methods are colored. c) An example segmentation for a cell with different methods, highlighting the advantage of automatic prompting for complex cell morphologies.

To address these questions we: (i) Extensively benchmark recent foundation models for general-purpose and microscopy segmentation. (ii) Develop a new instance segmentation algorithm that operates on top of (fine-tuned) SAMs without the need for additional training. Our results show substantial improvements in performance through our improved segmentation approach, competitive with the previous state-of-the-art. Further, they show that foundation models for microscopy have an edge over SAM3 and a substantial influence of the training data on model performance. See Fig. 1 for an overview.

## 2. Related Work

Vision foundation models (VFMs) have emerged after the success of large language models (LLMs) as generally capable language processors (Brown et al., 2020). VFMs can be divided into three categories: (i) vision and text models, such as CLIP (Radford et al., 2021) and SigLIP (Zhai et al., 2023), that are trained via contrastive learning on image-text-pairs. They are a key ingredient of multi-modal LLMs. (ii) general-purpose vision encoders such as Dino v2 (Oquab et al., 2023) and v3 (Siméoni et al., 2025) that are trained via self supervised learning, enabling diverse downstream tasks by fine-tuning small decoders. And (iii) foundation models for segmentation. Among those, SAM (Kirillov et al., 2023) is most popular. It supports interactive segmentation based on point or box prompts, and also supports automatic segmentation. SAM2 (Ravi et al., 2024) extends its capabilities to video segmentation. SAM3 (Carion et al., 2025) introduces “concept-based” segmentation, enabling prompting with short text descriptions or example images with object annotations. Other examples of segmentation foundation models include SegGPT (Wang et al., 2023), which supports example-based learning, LISA (Lai et al., 2023), which extends an LLM with segmentation capabilities, and SEEM (Zou et al., 2023), which offers a wide range of prompting options.

Many segmentation foundation models for biomedical images have emerged. They are predominantly based on SAM(2). Their exact adaptation strategies vary and can be divided into three categories, including combinations thereof:

1. By automatically deriving prompts (points or bounding boxes) for objects in the image and then using SAM to predict the corresponding masks (Chen et al., 2024).
2. By using SAM’s encoder as backbone of a model for automatic segmentation that is trained on domain-specific data (He et al., 2025).
3. By fine-tuning SAM’s architecture for promptable segmentation on domain-specific data (Ma et al., 2024; Archit et al., 2025b; Cheng et al., 2023).

Several segmentation foundation models have been proposed for medical imaging, e.g. CT, MRI, X-Ray, (see previous paragraph). Instance segmentation in microscopy is another important application. The most popular models are CellSAM (Israel et al., 2023) for light microscopy segmentation,  $\mu$ SAM (Archi et al., 2025a) for light and electron microscopy segmentation, its extension PathoSAM (Griebel et al., 2025) for histopathology, CellPoseSAM (Pachitariu et al., 2025) for light microscopy and histopathology, and CellViT (Hörst et al., 2024) for histopathology. See als Sec. 3.1.

## 3. Methods

We first review foundation models for microscopy segmentation (Sec. 3.1), then explain our new instance segmentation approach (Sec. 3.2) and evaluation methodology (Sec. 3.3).

### 3.1. Instance segmentation with SAM-based models

The models of the SAM family all support automatic instance segmentation. SAM (Kirillov et al., 2023) and SAM2 (Ravi et al., 2024) are trained with an objective for prompt-based

segmentation. They support automatic segmentation by covering the input image with a grid of point prompts, segmenting each prompt, and the merging the predicted masks via non-maximum suppression (NMS). This mode is called automatic mask generation (AMG). In contrast, SAM3 (Carion et al., 2025) predicts an instance segmentation directly with a DETR-style approach (Carion et al., 2020). All three models were primarily trained on natural images with segmentation annotations, 11 million images with 1 billion annotations for SAM, an additional 50k videos with 642k object track annotations for SAM2, and an additional 5 million images and 50k videos for SAM3.

SAM has been studied extensively for microscopy data (see e.g. (Archit et al., 2025a), Sec. 4). It yields good segmentation results for easy tasks (e.g. well separated nuclei) via AMG, but fails for more difficult tasks. AMG with SAM2 performs worse for microscopy (see Sec. 4). SAM3 has been published very recently. To our knowledge we are the first to evaluate it for microscopy.

While SAM itself fails at difficult microscopy segmentation tasks, the state-of-the-art models for microscopy are built on top of it, following one of the strategies 1.-3. outlined in Sec. 2. The simplest strategy (2) is to train a new decoder on top of SAM’s image encoder (initialized with its pretrained weights) that outputs an intermediate prediction, followed by (non-differentiable) post-processing to obtain instances. This approach is implemented by CellPose-SAM (Pachitariu et al., 2025), which chose the CellPose instance segmentation method (Stringer et al., 2021) and was trained on 22,826 light microscopy and histopathology images with 3.34 million annotated cells and nuclei, and Cell-ViT, which chose the HoverNet semantic instance segmentation method (Graham et al., 2019) and was trained on the PanNuke (Gamper et al., 2020), consisting of 200,000 annotated nuclei.

CellSAM (Israel et al., 2023) takes a more complex approach: it trains a bounding box detection decoder (CellFinder) on top of SAM’s decoder and then uses its predictions as box prompts for SAM’s prompt encoder to segment instance masks. The mask decoder is also finetuned, i.e., corresponding to a combination of strategies 1 and 2. CellSAM was trained on 10 datasets with annotated cells and nuclei.

$\mu$ SAM (Archit et al., 2025a) finetunes the entire SAM architecture for promptable segmentation while adding a decoder for instance segmentation that predicts foreground probabilities as well as normalized distances to object centers and boundaries. These predictions serve as input to a watershed for instance segmentation. The procedure is called automatic instance segmentation (AIS).  $\mu$ SAM trained on 17,000 light microscopy images with over 2 million annotated cells and nuclei; a different version of the model for electron microscopy also exists. PathoSAM (Griebel et al., 2025) replicates this effort for histopathology. It was trained on 5,000 images with 400,000 annotated nuclei. This approach corresponds to a combination of strategies 2 and 3.

### 3.2. Automatic prompt generation

We observe that none of the prior SAM-based microscopy foundation models combine all three adaptation strategies, i.e. none of them combine automatically derived prompts (1) with a custom decoder (2), and finetuning for promptable segmentation (3). While CellSAM comes close to this combination, it does not finetune for promptable segmentation and thus relies heavily on the box predictions, which are translated one-to-one to masks.

Hence, if an object is not correctly detected, it cannot be recovered by SAM’s promptable segmentation. This becomes more likely under a domain shift, leading to diminished generalization, which could be avoided with a more flexible prompting strategy. We also observe that segmentation based on prompts derived from annotated objects performs significantly better than fully automated segmentation (Archit et al., 2025a). Hence, a suitable automatic prompting strategy should be able to improve upon automatic instance segmentation without prompting, exemplified for a cell with complex morphology in Fig. 1.

We implement a new instance segmentation method called automatic prompt generation (APG) in the  $\mu$ SAM framework based on this hypothesis. It works as follows:

1. Apply the image encoder and segmentation decoder to predict foreground probabilities  $fg$  and normalized boundary as well as center distances,  $d_b$  and  $d_c$ .
2. Apply the thresholds  $t_{fg}$ ,  $t_b$  and  $t_c$  to  $fg$ ,  $d_b$ , and  $d_c$ , respectively.
3. Apply connected components to the intersection of the three binary masks from 2.
4. Derive a point prompt for each component by computing the maximum of the boundary distance transform per component.
5. Apply prompt encoder and mask decoder to these prompts to obtain mask and IoU predictions. The latter give a quality estimate for each predicted mask.
6. Apply a size filter  $s$  to the masks.
7. Compute the pairwise overlap of predicted masks and apply NMS with threshold  $t_{nms}$  based on predicted IoUs to filter overlapping masks.

Here, we set  $t_{fg} = 0.5$ ;  $t_b$ ,  $t_c$ ,  $s$ , and  $t_{nms}$  are parameters with default values 0.5, 0.5, 25, and 0.9. We run all experiments with these defaults. See Fig. 1 for a schematic overview of APG.

Note that step 2 and 3 are the same as in the AIS method of  $\mu$ SAM, which then uses the components as seeds for a watershed. However,  $t_b$  and  $t_c$  have to be determined such that each object is covered with a single component, leading to a trade-off between over- and under-segmentation. In APG, we can choose these values so that multiple prompts are derived for one object, then filtered in step 7 by NMS. Note that APG can be applied to the  $\mu$ SAM (and PathSAM) model as is without the need for retraining.

### 3.3. Datasets & metrics

We evaluate APG and other foundation models on 36 datasets from four different tasks and domains: nucleus segmentation in fluorescence microscopy (Moen et al., 2019; Ljosa et al., 2012; Arvidsson et al., 2023; Kromp et al., 2020; Vijayan et al., 2024; Zheng et al., 2023; Alwes et al., 2016; Caicedo et al., 2019; Mahbod et al., 2021b), cell segmentation in fluorescence microscopy (Stringer et al., 2021; Shi et al., 2025; Greenwald et al., 2022; Wolny et al., 2020; Pape et al., 2021; Ouyang et al., 2019; Willis et al., 2016; Bondarenko et al., 2023), cell segmentation in label-free microscopy (Edlund et al., 2021; Cutler et al., 2022; Spahn et al., 2022; Tsai et al., 2019; Vicar et al., 2021; Zargari et al., 2023; Seiffarth et al.,

2024; Gupta et al., 2023; Dietler et al., 2020), and nucleus segmentation in histopathology (Wang et al., 2024; Naji et al., 2024; Kumar et al., 2019; Gamper et al., 2020; Naylor et al., 2018; Mahbod et al., 2024; Schuiveling et al., 2025; Vadori et al., 2025; Mahbod et al., 2021a). We use the test splits for all of these datasets. For volumetric datasets we run and evaluate the segmentation in 2D and sub-sample the test sets for efficiency reasons. Note that some of the methods we evaluate were trained on some of these datasets (though not only on the train splits, not on the test splits). See Fig. 2 for details. We use the mean segmentation accuracy, following (Caicedo et al., 2019), to evaluate instance segmentation results. See App A for details.

## 4. Results

We evaluate SAM (w/ AMG), SAM3,  $\mu$ SAM (w/ AIS, APG), PathoSAM (w/ AIS, APG), CellPose-SAM, and CellSAM (see Sec. 3) on the data described in Sec. 3.3. We also evaluate CellPose3 (Stringer and Pachitariu, 2025), which uses a convolutional architecture and a smaller training set but is otherwise similar to CellPose-SAM. PathoSAM is only evaluated on histopathology data and  $\mu$ SAM is not evaluated in this domain to account for the specific focus of these models. Note that we do not evaluate SAM2 (w/ AMG) as we found it to be inferior to SAM in this setting. It did not segment any objects for several datasets in initial experiments we ran. SAM3 is prompted with a single short noun and then segments all images automatically, see Sec. 4.2 for details.

The result summary is shown in Fig. 1, the results for all datasets in Fig. 2, reported separately across the four data domains (see Sec. 3.3). We highlight the three top performing methods and indicate whether the model was trained on the respective data’s train split. APG improves the  $\mu$ SAM model compared to AIS for all label-free microscopy cases, including a very substantial improvement for TOIAM. It improves 3 / 9 datasets for cell segmentation in fluorescence microscopy, with only modest differences in segmentation quality, improves 7 / 9 datasets for nucleus segmentation in fluorescence, and 7 / 9 for nucleus segmentation in histopathology. Overall, we find that CellPoseSAM and APG are consistently among the best approaches, ranking among the top three methods for all four domains. We find that CellSAM performs worse than the other microscopy foundation models, but consistently better than SAM and SAM3.

### 4.1. Comparison of APG strategies

We also compare an alternative strategy for deriving prompts in APG, using the foreground-restricted maxima of the boundary distance predictions. The corresponding results are shown in Fig. 3, showing that our connected components based strategy is superior.

### 4.2. Prompting strategies for SAM3

The results for SAM3 in Fig. 2 were obtained by prompting the model with a short noun phrase describing the objects in the image. We tried different nouns for the respective segmentation tasks and found that SAM3 was sensitive to the choice of noun and not yet aware of domain-specific terminology. For example, it correctly segmented nuclei in light



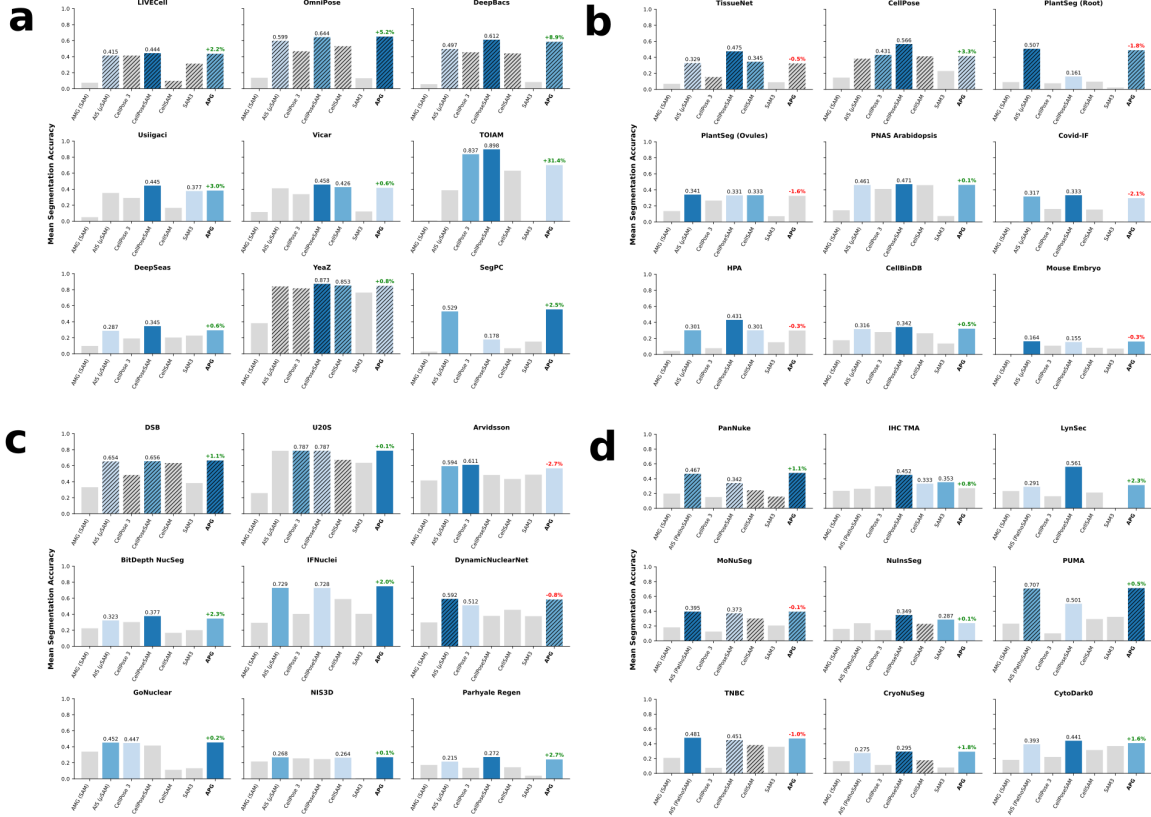


Figure 2: Results for 36 microscopy segmentation datasets in four different modalities: cells (fluorescence, a), cells (label-free, b), nuclei (fluorescence, c), and nuclei (histopathology, d). We indicate the top-3 ranked method with blue colors and methods that were trained on the corresponding training split with textured bars. For our method (APG) we indicate the absolute performance difference with respect to the reference method (AIS).

microscopy when prompted with “cells”, but not when prompted with “nuclei”. We did not yet prompt SAM3 with an example image and representative segmented objects.

## 5. Discussion

We introduced APG, a new method for instance segmentation with SAM-based foundation models. It consistently and substantially improved the results of  $\mu$ SAM and PathoSAM compared to the prior AIS approach and is competitive with the prior state-of-the-art CellPoseSAM. APG was applied without the need for re-training. It could also be applied to other SAM-based models, e.g. CellPoseSAM. However, in this case it would either rely on another model for prompt-based segmentation or joint training of CellPoseSAM with the objective for interactive segmentation.

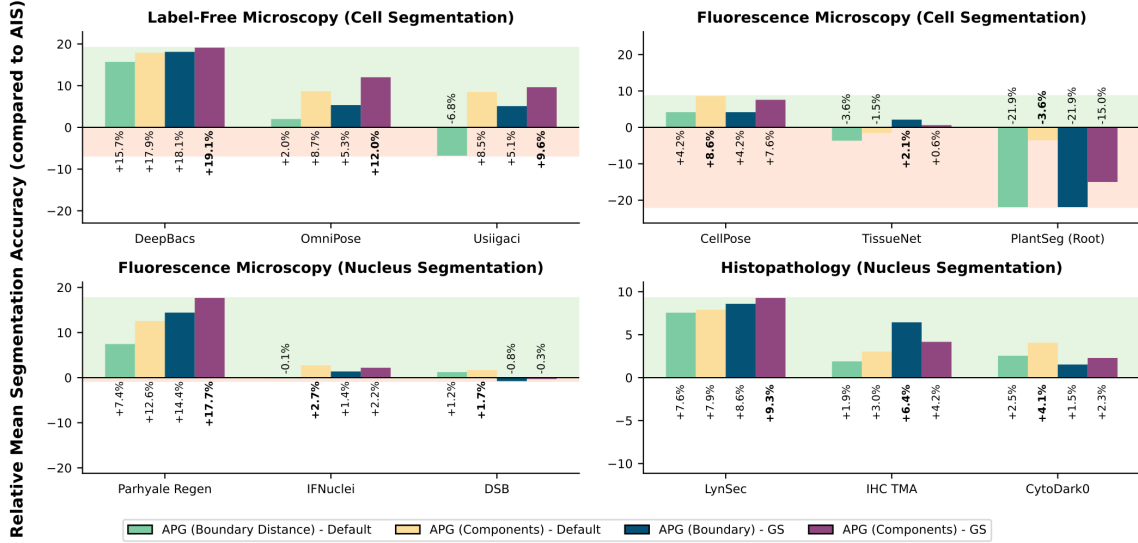


Figure 3: Comparison of our connected component based APG (APG (Components)) strategy with an alternative deriving prompts from distance map maxima (APG (Boundary)) for the four different modalities (a-d), reporting the relative mean segmentation accuracy difference w.r.t AIS.

Furthermore, we evaluated SAM3, the latest mode of the SAM family, for microscopy. It performed well – though not yet competitive with domain specific foundation models –, despite being trained on only two relevant datasets (Edlund et al., 2021; Gamper et al., 2020). Finetuning SAM3 on microscopy data promises to yield a further improvements in this domain. We did not yet test prompting SAM3 with examples of annotated images, which could lead to substantial improvements without further training.

Overall, we saw a clear impact of model training data on performance, with models typically performing similar when trained on the respective data’s training split, and more pronounced for “out-of-domain” datasets. Yet, we found that APG can lead to substantial improvements for the in-domain (e.g. OmniPose, DeepBacs) and out-of-domain (e.g. TOIAM) case. Overall, the performance of models seems to correlate with the size of their domain-specific training data, CellPose and  $\mu$ SAM (w/ AIS & APG) having the largest training sets and overall best performances. This observation also indicates that further finetuning to improve for specific applications remains relevant, as already demonstrated in prior work (Teuber et al., 2025; Zhou et al., 2024) that demonstrated substantial improvements through training on small datasets – even a single image. These approaches would also directly translate to specifically improving APG.

Furthermore, APG could likely be improved by incorporating box prompts, which generally perform better than point prompts (Archit et al., 2025a). We performed initial experiments to derive candidate box prompts from the  $\mu$ SAM decoder predictions. However, we found that deriving a set of high-quality prompts that over-sample objects was challeng-



ing and did not yet find a strategy competitive with the simpler point prompt derivation. Future work, such as iterative prompt derivation, may enable such improvements.

Finally, a limitation of our study is the restriction to 2D evaluation, also for 3D datasets. CellPoseSAM,  $\mu$ SAM, and SAM3 support 3D segmentation, so this evaluation would provide a further valuable comparison and would further inform how to improve microscopy foundation models. We plan to address this problem in future work, potentially including adaptations of SAM2 and/or SAM3 to microscopy.

## Acknowledgments

The work of Anwai Archit was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) - PA 4341/2-1. Constantin Pape is supported by the German Research Foundation (Deutsche Forschungsgemeinschaft, DFG) under Germany’s Excellence Strategy - EXC 2067/1-390729940. This work is supported by the Ministry of Science and Culture of Lower Saxony through funds from the program zukunft.niedersachsen of the Volkswagen Foundation for the ‘CAIMed – Lower Saxony Center for Artificial Intelligence and Causal Methods in Medicine’ project (grant no. ZN4257). It was also supported by the Google Research Scholarship “Vision Foundation Models for Bioimage Segmentation”. We gratefully acknowledge the computing time granted by the Resource Allocation Board and provided on the supercomputer Emmy at NHR@Göttingen as part of the NHR infrastructure, under the project nim00007.

## References

- Frederike Alwes, Camille Enjolras, and Michalis Averof. Live imaging reveals the progenitors and cell dynamics of limb regeneration. *Elife*, 5:e19766, 2016.
- Anwai Archit, Luca Freckmann, Sushmita Nair, Nabeel Khalid, Paul Hilt, Vikas Rajashekar, Marei Freitag, Carolin Teuber, Melanie Spitzner, Constanza Tapia Contreras, et al. Segment anything for microscopy. *Nature Methods*, 22(3):579–591, 2025a.
- Anwai Archit, Luca Freckmann, and Constantin Pape. Medicosam: Towards foundation models for medical image segmentation. *arXiv preprint arXiv:2501.11734*, 2025b.
- Malou Arvidsson, Salma Kazemi Rashed, and Sonja Aits. An annotated high-content fluorescence microscopy dataset with hoechst 33342-stained nuclei and manually labelled outlines. *Data in Brief*, 46:108769, 2023.
- Stuart Berg, Dominik Kutra, Thorben Kroeger, Christoph N Straehle, Bernhard X Kausler, Carsten Haubold, Martin Schiegg, Janez Ales, Thorsten Beier, Markus Rudy, et al. Ilastik: interactive machine learning for (bio) image analysis. *Nature methods*, 16(12):1226–1232, 2019.
- Vladyslav Bondarenko, Mikhail Nikolaev, Dimitri Kromm, Roman Belousov, Adrian Wolny, Marloes Blotenburg, Peter Zeller, Saba Rezakhani, Johannes Hugger, Virginie Uhlmann, Lars Hufnagel, Anna Kreshuk, Jan Ellenberg, Alexander van Oudenaarden, Anna Erzberger, Matthias P Lutolf, and Takashi Hiiragi. Embryo-uterine interaction coordinates mouse embryogenesis during implantation. *The EMBO Journal*, 42(17), July 2023. ISSN 1460-2075. doi: 10.15252/emboj.2022113280. URL <http://doi.org/10.15252/emboj.2022113280>.
- Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, et al. Language models are few-shot learners. *Advances in neural information processing systems*, 33:1877–1901, 2020.

- Juan C Caicedo, Allen Goodman, Kyle W Karhohs, Beth A Cimini, Jeanelle Ackerman, Marzieh Haghighi, CherKeng Heng, Tim Becker, Minh Doan, Claire McQuin, et al. Nucleus segmentation across imaging experiments: the 2018 data science bowl. *Nature methods*, 16(12):1247–1253, 2019.
- Nicolas Carion, Francisco Massa, Gabriel Synnaeve, Nicolas Usunier, Alexander Kirillov, and Sergey Zagoruyko. End-to-end object detection with transformers. In *European conference on computer vision*, pages 213–229. Springer, 2020.
- Nicolas Carion, Laura Gustafson, Yuan-Ting Hu, Shoubhik Debnath, Ronghang Hu, Didac Suris, Chaitanya Ryali, Kalyan Vasudev Alwala, Haitham Khedr, Andrew Huang, et al. Sam 3: Segment anything with concepts. *arXiv preprint arXiv:2511.16719*, 2025.
- Srinivas Niranj Chandrasekaran, Beth A Cimini, Amy Goodale, Lisa Miller, Maria Kost-Alimova, Nasim Jamali, John G Doench, Briana Fritchman, Adam Skepner, Michelle Melanson, et al. Three million images and morphological profiles of cells treated with matched chemical and genetic perturbations. *Nature Methods*, 21(6):1114–1121, 2024.
- Zhen Chen, Qing Xu, Xinyu Liu, and Yixuan Yuan. Un-sam: Universal prompt-free segmentation for generalized nuclei images. *arXiv preprint arXiv:2402.16663*, 2024.
- Junlong Cheng, J Ye, Z Deng, J Chen, T Li, H Wang, Y Su, Z Huang, J Chen, L Jiang, et al. Sam-med2d. *arXiv preprint arXiv:2308.16184*, 2023.
- Ryan Conrad and Kedar Narayan. Instance segmentation of mitochondria in electron microscopy images with a generalist deep learning model trained on a diverse dataset. *Cell Systems*, 14(1):58–71, 2023.
- Kevin J Cutler, Carsen Stringer, Teresa W Lo, Luca Rappez, Nicholas Stroustrup, S Brook Peterson, Paul A Wiggins, and Joseph D Mougous. Omnipose: a high-precision morphology-independent solution for bacterial cell segmentation. *Nature methods*, 19(11):1438–1448, 2022.
- Nicola Dietler, Matthias Minder, Vojislav Gligorovski, Augoustina Maria Economou, Denis Alain Henri Lucien Joly, Ahmad Sadeghi, Chun Hei Michael Chan, Mateusz Koziński, Martin Weigert, Anne-Florence Bitbol, and Sahand Jamal Rahi. A convolutional neural network segments yeast microscopy images with high accuracy. *Nature Communications*, 11(1), November 2020. ISSN 2041-1723. doi: 10.1038/s41467-020-19557-4. URL <http://doi.org/10.1038/s41467-020-19557-4>.
- Christoffer Edlund, Timothy R Jackson, Nabeel Khalid, Nicola Bevan, Timothy Dale, Andreas Dengel, Sheraz Ahmed, Johan Trygg, and Rickard Sjögren. Livecell—a large-scale dataset for label-free live cell segmentation. *Nature methods*, 18(9):1038–1045, 2021.
- Jevgenij Gamper, Navid Alemi Koohbanani, Ksenija Benes, Simon Graham, Mostafa Jahanifar, Syed Ali Khurram, Ayesha Azam, Katherine Hewitt, and Nasir Rajpoot. Pannuke dataset extension, insights and baselines. *arXiv preprint arXiv:2003.10778*, 2020.

- Estibaliz Gómez-de Mariscal, Carlos García-López-de Haro, Wei Ouyang, Laurène Donati, Emma Lundberg, Michael Unser, Arrate Muñoz-Barrutia, and Daniel Sage. Deepimagej: A user-friendly environment to run deep learning models in imagej. *Nature methods*, 18(10):1192–1195, 2021.
- Simon Graham, Quoc Dang Vu, Shan E Ahmed Raza, Ayesha Azam, Yee Wah Tsang, Jin Tae Kwak, and Nasir Rajpoot. Hover-net: Simultaneous segmentation and classification of nuclei in multi-tissue histology images. *Medical image analysis*, 58:101563, 2019.
- Noah F Greenwald, Geneva Miller, Erick Moen, Alex Kong, Adam Kagel, Thomas Dougherty, Christine Camacho Fullaway, Brianna J McIntosh, Ke Xuan Leow, Morgan Sarah Schwartz, et al. Whole-cell segmentation of tissue images with human-level performance using large-scale data annotation and deep learning. *Nature biotechnology*, 40(4):555–565, 2022.
- Titus Griebel, Anwai Archit, and Constantin Pape. Segment anything for histopathology. *arXiv preprint arXiv:2502.00408*, 2025.
- Anubha Gupta, Shiv Gehlot, Shubham Goswami, Sachin Motwani, Ritu Gupta, Álvaro García Faura, Dejan Štepec, Tomaž Martinčič, Reza Azad, Dorit Merhof, Afshin Bozorgpour, Babak Azad, Alaa Sulaiman, Deepanshu Pandey, Pradyumna Gupta, Sumit Bhattacharya, Aman Sinha, Rohit Agarwal, Xinyun Qiu, Yucheng Zhang, Ming Fan, Yoonbeom Park, Daehong Lee, Joon Sik Park, Kwangyeol Lee, and Jaehyung Ye. Segpc-2021: A challenge amp; dataset on segmentation of multiple myeloma plasma cells from microscopic images. *Medical Image Analysis*, 83:102677, January 2023. ISSN 1361-8415. doi: 10.1016/j.media.2022.102677. URL <http://doi.org/10.1016/j.media.2022.102677>.
- Yufan He, Pengfei Guo, Yucheng Tang, Andriy Myronenko, Vishwesh Nath, Ziyue Xu, Dong Yang, Can Zhao, Benjamin Simon, Mason Belue, et al. Vista3d: A unified segmentation foundation model for 3d medical imaging. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pages 20863–20873, 2025.
- Fabian Hörst, Moritz Rempe, Lukas Heine, Constantin Seibold, Julius Keyl, Giulia Baldini, Selma Ugurel, Jens Siveke, Barbara Grünwald, Jan Egger, et al. Cellvit: Vision transformers for precise cell segmentation and classification. *Medical Image Analysis*, 94:103143, 2024.
- Uriah Israel, Markus Marks, Rohit Dilip, Qilin Li, Morgan Sarah Schwartz, Elora Pradhan, Edward Pao, Shenyi Li, Alexander Pearson-Goulart, Pietro Perona, et al. A foundation model for cell segmentation.  *biorxiv*, 2023.
- Alexander Kirillov, Eric Mintun, Nikhila Ravi, Hanzi Mao, Chloe Rolland, Laura Gustafson, Tete Xiao, Spencer Whitehead, Alexander C Berg, Wan-Yen Lo, et al. Segment anything. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 4015–4026, 2023.

- Florian Kromp, Eva Bozsaky, Fikret Rifatbegovic, Lukas Fischer, Magdalena Ambros, Maria Berneder, Tamara Weiss, Daria Lazic, Wolfgang Dörr, Allan Hanbury, et al. An annotated fluorescence image dataset for training nuclear segmentation methods. *Scientific Data*, 7(1):262, 2020.
- Neeraj Kumar, Ruchika Verma, Deepak Anand, Yanning Zhou, Omer Fahri Onder, Efstratios Tsougenis, Hao Chen, Pheng-Ann Heng, Jiahui Li, Zhiqiang Hu, et al. A multi-organ nucleus segmentation challenge. *IEEE transactions on medical imaging*, 39(5):1380–1391, 2019.
- X Lai, Z Tian, Y Chen, et al. Lisa: Reasoning segmentation via large language model. *arXiv preprint arXiv: 230800692*, 2023.
- Merlin Lange, Alejandro Granados, Shruthi VijayKumar, Jordão Bragantini, Sarah Ancheta, Yang-Joon Kim, Sreejith Santhosh, Michael Borja, Hirofumi Kobayashi, Erin McGeever, et al. A multimodal zebrafish developmental atlas reveals the state-transition dynamics of late-vertebrate pluripotent axial progenitors. *Cell*, 187(23):6742–6759, 2024.
- Vebjorn Ljosa, Katherine L Sokolnicki, and Anne E Carpenter. Annotated high-throughput microscopy image sets for validation. *Nature methods*, 9(7):637, 2012.
- Jun Ma, Yuting He, Feifei Li, Lin Han, Chenyu You, and Bo Wang. Segment anything in medical images. *Nature Communications*, 15(1):654, 2024.
- Amirreza Mahbod, Gerald Schaefer, Benjamin Bancher, Christine Löw, Georg Dorffner, Rupert Ecker, and Isabella Ellinger. Cryonuseg: A dataset for nuclei instance segmentation of cryosectioned hamp;e-stained histological images. *Computers in Biology and Medicine*, 132:104349, May 2021a. ISSN 0010-4825. doi: 10.1016/j.compbimed.2021.104349. URL <http://dx.doi.org/10.1016/j.compbimed.2021.104349>.
- Amirreza Mahbod, Gerald Schaefer, Christine Löw, Georg Dorffner, Rupert Ecker, and Isabella Ellinger. Investigating the impact of the bit depth of fluorescence-stained images on the performance of deep learning-based nuclei instance segmentation. *Diagnostics*, 11(6):967, 2021b.
- Amirreza Mahbod, Christine Polak, Katharina Feldmann, Rumsha Khan, Katharina Gelles, Georg Dorffner, Ramona Woitek, Sepideh Hatamikia, and Isabella Ellinger. Nuinsseg: a fully annotated dataset for nuclei instance segmentation in h&e-stained histological images. *Scientific Data*, 11(1):295, 2024.
- Erick Moen, Enrico Borba, Geneva Miller, Morgan Schwartz, Dylan Bannon, Nora Koe, Isabella Camplisson, Daniel Kyme, Cole Pavelchek, Tyler Price, et al. Accurate cell tracking and lineage construction in live-cell imaging experiments with deep learning. *Biorxiv*, page 803205, 2019.
- Sarah Muth, Frederieke Moschref, Luca Freckmann, Sophia Mutschall, Ines Hojas-Garcia-Plaza, Julius N Bahr, Arsen Petrovic, Thanh Thao Do, Valentin Schwarze, Anwai Archit, et al. Synapsenet: deep learning for automatic synapse reconstruction. *Molecular Biology of the Cell*, 36(10):ar127, 2025.

- Hussein Naji, Lucas Sancere, Adrian Simon, Reinhard Büttner, Marie-Lisa Eich, Philipp Lohneis, and Katarzyna Bożek. Holy-net: Segmentation of histological images of diffuse large b-cell lymphoma. *Computers in Biology and Medicine*, 170:107978, 2024.
- Peter Naylor, Marick Laé, Fabien Reyat, and Thomas Walter. Segmentation of nuclei in histopathology images by deep regression of the distance map. *IEEE transactions on medical imaging*, 38(2):448–459, 2018.
- Maxime Oquab, Timothée Darcet, Théo Moutakanni, Huy Vo, Marc Szafraniec, Vasil Khalidov, Pierre Fernandez, Daniel Haziza, Francisco Massa, Alaaeldin El-Nouby, et al. Dinov2: Learning robust visual features without supervision. *arXiv preprint arXiv:2304.07193*, 2023.
- Wei Ouyang, Casper F Winsnes, Martin Hjelmare, Anthony J Cesnik, Lovisa Åkesson, Hao Xu, Devin P Sullivan, Shubin Dai, Jun Lan, Park Jinmo, et al. Analysis of the human protein atlas image classification competition. *Nature methods*, 16(12):1254–1261, 2019.
- Wei Ouyang, Fynn Beuttenmueller, Estibaliz Gómez-de Mariscal, Constantin Pape, Tom Burke, Carlos Garcia-López-de Haro, Craig Russell, Lucía Moya-Sans, Cristina De-La-Torre-Gutiérrez, Deborah Schmidt, et al. Bioimage model zoo: a community-driven resource for accessible deep learning in bioimage analysis. *BioRxiv*, pages 2022–06, 2022.
- Marius Pachitariu, Michael Rariden, and Carsen Stringer. Cellpose-sam: superhuman generalization for cellular segmentation. *bioRxiv*, pages 2025–04, 2025.
- Constantin Pape, Roman Remme, Adrian Wolny, Sylvia Olberg, Steffen Wolf, Lorenzo Cerone, Mirko Cortese, Severina Klaus, Bojana Lucic, Stephanie Ullrich, et al. Microscopy-based assay for semi-quantitative detection of sars-cov-2 specific antibodies in human sera: A semi-quantitative, high throughput, microscopy-based assay expands existing approaches to measure sars-cov-2 specific antibody levels in human sera. *Bioessays*, 43(3): 2000257, 2021.
- Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, et al. Learning transferable visual models from natural language supervision. In *International conference on machine learning*, pages 8748–8763. PmLR, 2021.
- Nikhila Ravi, Valentin Gabeur, Yuan-Ting Hu, Ronghang Hu, Chaitanya Ryali, Tengyu Ma, Haitham Khedr, Roman Rädle, Chloe Rolland, Laura Gustafson, et al. Sam 2: Segment anything in images and videos. *arXiv preprint arXiv:2408.00714*, 2024.
- Uwe Schmidt, Martin Weigert, Coleman Broaddus, and Gene Myers. Cell detection with star-convex polygons. In *International conference on medical image computing and computer-assisted intervention*, pages 265–273. Springer, 2018.
- Mark Schuiveling, Hong Liu, Daniel Eek, Gerben E Breimer, Karijn PM Suijkerbuijk, Willeke AM Blokkx, and Mitko Veta. A novel dataset for nuclei and tissue segmentation in melanoma with baseline nuclei segmentation and tissue segmentation benchmarks. *GigaScience*, 14:giaf011, 2025.



- Johannes Seiffarth, Luisa Blöbaum, Richard D Paul, Nils Friederich, Angelo Jovin Yamachui Sitcheu, Ralf Mikut, Hanno Scharr, Alexander Grünberger, and Katharina Nöh. Tracking one-in-a-million: Large-scale benchmark for microbial single-cell tracking with experiment-aware robustness metrics. In *European Conference on Computer Vision*, pages 318–334. Springer, 2024.
- Can Shi, Jinghong Fan, Zhonghan Deng, Huanlin Liu, Qiang Kang, Yumei Li, Jing Guo, Jingwen Wang, Jinjiang Gong, Sha Liao, et al. Cellbindb: a large-scale multimodal annotated dataset for cell segmentation with benchmarking of universal models. *GigaScience*, 14:giaf069, 2025.
- Oriane Siméoni, Huy V Vo, Maximilian Seitzer, Federico Baldassarre, Maxime Oquab, et al. Dinov3. *arXiv preprint arXiv:2508.10104*, 2(4):5, 2025.
- Christoph Spahn, Estibaliz Gómez-de Mariscal, Romain F Laine, Pedro M Pereira, Lucas von Chamier, Mia Conduit, Mariana G Pinho, Guillaume Jacquemet, Séamus Holden, Mike Heilemann, et al. Deepbacs for multi-task bacterial image analysis using open-source deep learning approaches. *Communications Biology*, 5(1):688, 2022.
- Carsen Stringer and Marius Pachitariu. Cellpose3: one-click image restoration for improved cellular segmentation. *Nature methods*, 22(3):592–599, 2025.
- Carsen Stringer, Tim Wang, Michalis Michaelos, and Marius Pachitariu. Cellpose: a generalist algorithm for cellular segmentation. *Nature methods*, 18(1):100–106, 2021.
- Carolyn Teuber, Anwai Archit, and Constantin Pape. Parameter efficient fine-tuning of segment anything model for biomedical imaging. *arXiv preprint arXiv:2502.00418*, 2025.
- Hsieh-Fu Tsai, Joanna Gajda, Tyler FW Sloan, Andrei Rares, and Amy Q Shen. Usiigaci: Instance-aware cell tracking in stain-free phase contrast microscopy enabled by machine learning. *SoftwareX*, 9:230–237, 2019.
- Valentina Vadori, Jean-Marie Graïc, Antonella Peruffo, Giulia Vadori, Livio Finos, and Enrico Grisan. Cisca and cytodark0: a cell instance segmentation and classification method for histo (patho) logical image analyses and a new, open, nissl-stained dataset for brain cytoarchitecture studies. *Computers in Biology and Medicine*, 197:111018, 2025.
- Tomas Vicar, Jiri Chmelik, Roman Jakubicek, Larisa Chmelikova, Jaromir Gumulec, Jan Balvan, Ivo Provaznik, and Radim Kolar. Self-supervised pretraining for transferable quantitative phase image cell segmentation. *Biomedical optics express*, 12(10):6514–6528, 2021.
- Athul Vijayan, Tejasvinee Atul Mody, Qin Yu, Adrian Wolny, Lorenzo Cerrone, Soeren Strauss, Miltos Tsiantis, Richard S Smith, Fred A Hamprecht, Anna Kreshuk, et al. A deep learning-based toolkit for 3d nuclei segmentation and quantitative analysis in cellular and tissue context. *Development*, 151(14):dev202800, 2024.
- Ranran Wang, Yusong Qiu, Xinyu Hao, Shan Jin, Junxiu Gao, Heng Qi, Qi Xu, Yong Zhang, and Hongming Xu. Simultaneously segmenting and classifying cell nuclei by

- using multi-task learning in multiplex immunohistochemical tissue microarray sections. *Biomedical Signal Processing and Control*, 93:106143, 2024.
- Xinlong Wang, Xiaosong Zhang, Yue Cao, Wen Wang, Chunhua Shen, and Tiejun Huang. Seggpt: Segmenting everything in context. *arXiv preprint arXiv:2304.03284*, 2023.
- Lisa Willis, Yassin Refahi, Raymond Wightman, Benoit Landrein, José Teles, Kerwyn Casey Huang, Elliot M Meyerowitz, and Henrik Jönsson. Cell size and growth regulation in the arabidopsis thaliana apical stem cell niche. *Proceedings of the National Academy of Sciences*, 113(51):E8238–E8246, 2016.
- Adrian Wolny, Lorenzo Cerrone, Athul Vijayan, Rachele Tofanelli, Amaya Vilches Barro, Marion Louveaux, Christian Wenzl, Sören Strauss, David Wilson-Sánchez, Rena Lymbouridou, et al. Accurate and versatile 3d segmentation of plant tissues at cellular resolution. *Elife*, 9:e57613, 2020.
- Abolfazl Zargari, Gerrald A Lodewijk, Najmeh Mashhadi, Nathan Cook, Celine W Neudorf, Kimiasadat Araghbidikashani, Robert Hays, Sayaka Kozuki, Stefany Rubio, Eva Hrabeta-Robinson, et al. Deepsea is an efficient deep-learning model for single-cell segmentation and tracking in time-lapse microscopy. *Cell Reports Methods*, 3(6), 2023.
- Xiaohua Zhai, Basil Mustafa, Alexander Kolesnikov, and Lucas Beyer. Sigmoid loss for language image pre-training. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 11975–11986, 2023.
- Wei Zheng, Cheng Peng, Zeyuan Hou, Boyu Lyu, Mengfan Wang, Xuelong Mi, Shuoxuan Qiao, Yinan Wan, and Guoqiang Yu. Nis3d: a completely annotated benchmark for dense 3d nuclei image segmentation. *Advances in neural information processing systems*, 36:4741–4752, 2023.
- Peilin Zhou, Bo Du, and Yongchao Xu. Cellseg1: Robust cell segmentation with one training image. *arXiv preprint arXiv:2412.01410*, 2024.
- Xueyan Zou, Jianwei Yang, Hao Zhang, Feng Li, Linjie Li, Jianfeng Wang, Lijuan Wang, Jianfeng Gao, and Yong Jae Lee. Segment everything everywhere all at once. *Advances in neural information processing systems*, 36:19769–19782, 2023.

## Appendix A. Mean segmentation accuracy

The mean segmentation accuracy (mSA) is based on true positives ( $TP$ ), false negatives ( $FN$ ), and false positives ( $FP$ ), which are derived from the intersection over union (IoU) of predicted and true objects. Specifically,  $TP(t)$  is defined as the number of matches between predicted and true objects with an IoU above the threshold  $t$ ,  $FP(t)$  correspond to the number of predicted objects minus  $TP(t)$ , and  $FN(t)$  to the number of true objects minus  $TP(t)$ . The mean segmentation accuracy is computed over multiple thresholds:

$$\text{Mean Segmentation Accuracy} = \frac{1}{|\# \text{ thresholds}|} \sum_t \frac{TP(t)}{TP(t) + FP(t) + FN(t)}.$$

Here, we use thresholds  $t \in [0.5, 0.55, 0.6, 0.65, 0.7, 0.75, 0.8, 0.85, 0.9, 0.95]$ . For each dataset, we report the average mean segmentation accuracy over images in the test set.