PACE: Point Annotation-Based Cell Segmentation for Efficient Microscopic Image Analysis

Nabeel Khalid[1], Tiago Comassetto Froes[1], Maria Caroprese[5], Gillian Lovell[4], Johan Trygg[3,6], Andreas Dengel[1,2], and Sheraz Ahmed[1]

[1] German Research Center for Artificial Intelligence (DFKI) GmbH, Kaiserslautern 67663, Germany
firstname.lastname@dfki.de
[2] RPTU Kaiserslautern–Landau, Kaiserslautern 67663, Germany
[3] Sartorius Corporate Research, Sweden
[4] Sartorius, Corporate Research, Royston, United Kingdom
[5] Sartorius, BioAnalytics, Royston, United Kingdom
firstname.lastname@sartorius.com
[6] Computational Life Science Cluster (CLiC), Umeå University, Sweden

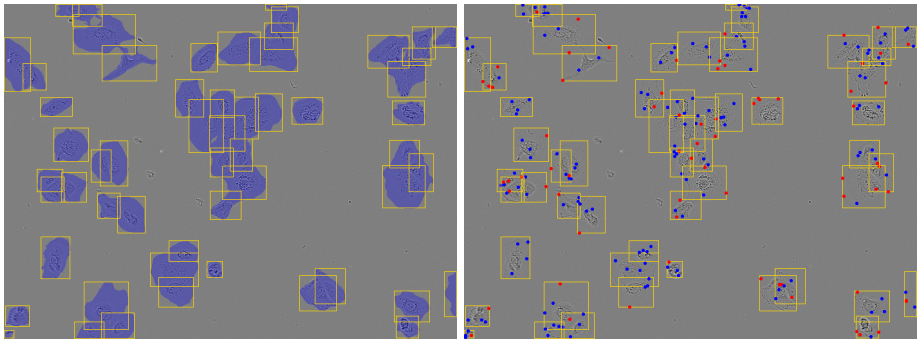
Abstract. Cells are essential to life because they provide the functional, genetic, and communication mechanisms essential for the proper functioning of living organisms. Cell segmentation is pivotal for any biological hypothesis validation/analysis i.e., to get valuable insights into cell behavior, function, diagnosis, and treatment. Deep learning-based segmentation methods have high segmentation precision, however, need fully annotated segmentation masks for each cell annotated manually by the experts, which is very laborious and costly. Many approaches have been developed in the past to reduce the effort required to annotate the data manually and even though these approaches produce good results, there is still a noticeable difference in performance when compared to fully supervised methods. To fill that gap, a weakly supervised approach, PACE, is presented, which uses only the point annotations and the bounding box for each cell to perform cell instance segmentation. The proposed approach not only achieves 99.8% of the fully supervised performance, but it also surpasses the previous state-of-the-art by a margin of more than 4%.

Keywords: cell segmentation · weakly supervised · point annotation · deep learning.

1 Introduction

Cells are the building blocks that make up all living organisms, from uncomplicated single-celled bacteria to complex multi-cellular organisms like humans. They provide the functional, genetic, and communication mechanisms essential for the proper functioning of living organisms. Cell segmentation is a key tool

for studying numerous aspects of cellular biology, and it allows researchers to study cell migration, cell differentiation, cell proliferation, cell physiology, gene expression patterns, and cell-cell communication in detail. Over the last decade, significant progress in deep learning-based (DL) approaches [15, 4, 7–9, 14, 16] for cell segmentation has been achieved. In a fully supervised setting, DL approaches require fully annotated data for training, with the boundary of each cell defined by the field experts. Manually defining the boundary of each cell in the microscopic images is very laborious and costly. In the natural image datasets like COCO [12], it takes an average of 79.2 seconds to draw a full mask for each object whereas the bounding box for each object takes only 7 seconds, which makes it 11 times faster than annotating the boundary of each object [13].



(a) Fully Supervised (b) Weakly Supervised (4-point)

Fig. 1: **Fully supervised (a) vs. weakly supervised (b)** example. The fully supervised method needs a full mask, whereas the presented weakly supervised approach, PACE, needs only the bounding box and the point annotations. The blue and red points represent whether the point lies on the cell or outside, respectively.

In the microscopic image analysis domain, the LIVECell dataset [4] is among the largest and most comprehensive datasets in cell biology research. It contains more than 1.6 million cells with an average cell density per image higher than any other publically available datasets in the cell biology research domain i.e, 313, which is almost 55 times more than the EVICAN[14] dataset. Annotating cells in microscopic images is more challenging than annotating objects in natural images due to the smaller scale, higher complexity, greater variability, and higher degree of noise in the images. Manually annotating the boundary of each cell in the image in the LIVECell dataset takes 46 seconds on average. Mask annotation time and complexity depend on cell culture morphology and density. Cell culture BV2 contains up to three thousand cells in some images, packed densely together, which makes it very hard for even the experts to identify the boundaries of the cells.

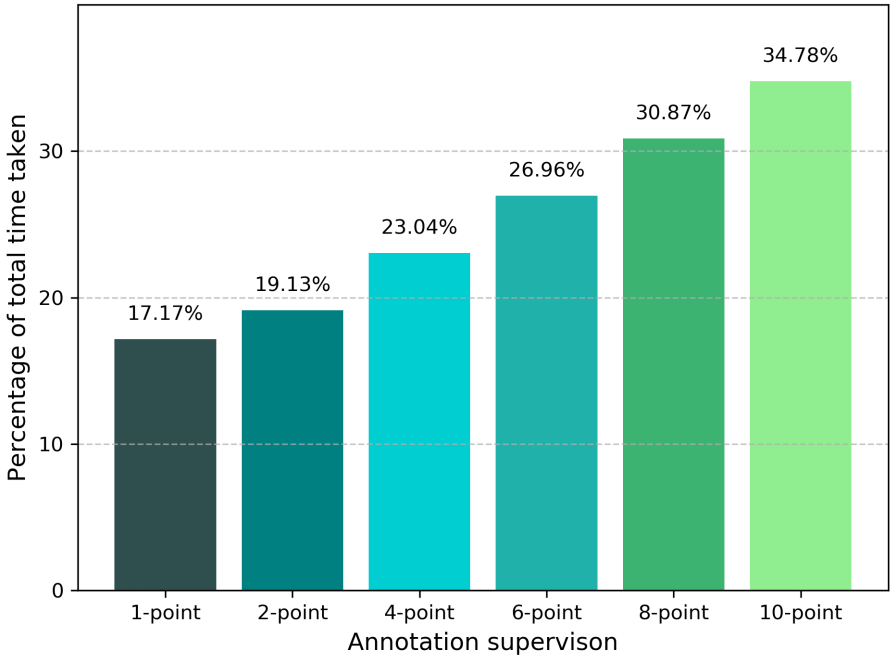


Fig. 2: **Annotation time for different point-supervision on the LIVECell dataset.** Labeling as many as 4 points per cell instance instead of the fully supervised (segmentation mask) annotation takes 23.04% of the total time spent on annotating the full mask for each cell and is 4.34x faster.

There is a significant amount of unlabeled cellular data available in the cell biology domain that has not been annotated for cell instance segmentation. Without annotations, this data cannot be used to train supervised DL models for cell instance segmentation. This means that the full potential of the data cannot be realized, as it is not being used to improve the accuracy and efficiency of cell segmentation algorithms. That is the reason why there is a need for a weakly supervised approach to perform cell segmentation and minimize the time and expert knowledge required in the annotation of data for the fully supervised methods. To address this issue, this paper presents a weakly supervised approach for cell segmentation, PACE, which requires only the bounding box and point annotations inside the bounding box for each cell to perform the task of cell segmentation. Figure 1 demonstrates the difference in the annotation required for the fully supervised methods 1a and the proposed weakly supervised method 1b. For the proposed weakly supervised, the first step is drawing the bounding boxes which take around $\sim$ 7 seconds per cell. After that random points are generated automatically inside the bounding boxes and the annotator only has to identify whether the point lies on the cell or outside, which takes $\sim$ 0.9

seconds. Fig. 2 provides insights into the annotation time required for different point supervision methods compared to the fully supervised method. Considering just one point for training saves us more than 82% (5.2x faster) of the total time required in labeling the data for the fully supervised method. Similarly, 23.04% (4.34x faster) and 30.87% (3.24x faster) of the total time spent on the annotation of the fully supervised method is needed for 4- and 8-points respectively. The main contributions of this study are as follows:

1. An end-to-end pipeline for weakly supervised point-based cell segmentation, PACE, using Cascade Mask R-CNN [1], Feature pyramid Network [11] with ResNeSt-200 [17], and bilinear interpolation [3].
2. Evaluation of the proposed approach using different point labels to examine the impact on the performance. Achieved 99.8% of the fully supervised performance using PACE with 8-point labels with a significant reduction in the time required for data annotation.
3. Outperformed the state-of-the-art method, Point2Mask [10], by a margin of 4.3%.

2 Related Work

Many different deep learning-based approaches [8, 7, 9, 16, 14] have been developed using the EVICAN [14] and the LIVECell dataset[4]. The Anchor-based method reported in the LIVECell paper achieved 47.89% mask mAP. However, the annotations required for training deep learning models are often time-consuming and challenging to obtain. To address this issue, weakly supervised or semi-supervised learning approaches have been proposed to reduce the annotation burden. Weakly supervised approaches like image tags [18], points [2], and missing annotations [5] have been proposed.
Khalid et al. (2022) [10] proposed Point2Mask, an approach for cell segmentation using the bounding box and the points instead of the full mask. Point2Mask achieved 99.2% (43.53%) of the fully supervised performance (43.90%) using just 6-point labels, saving more than 70% of the time required in annotating the full masks for the cells in the LIVECell dataset. Point2Mask used Mask R-CNN with ResNet-50.

3 PACE: The Proposed Approach

Fig. 3 provides a system overview of PACE. The proposed method is based on Cascade Mask R-CNN [1], Feature Pyramid Network [12], ResNeSt-200 [17] and Deformable Convolution. The proposed pipeline is composed of three blocks.

3.1 Backbone

The purpose of the backbone in the proposed method is to extract feature maps from the input image at different scales. The backbone is composed of Feature

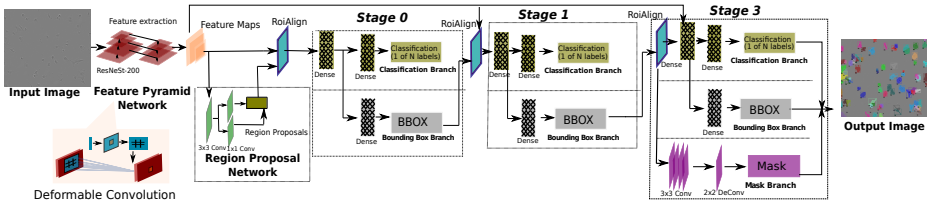


Fig. 3: **System overview of the PACE pipeline for weakly supervised cell segmentation.** The input image is passed to the proposed pipeline and the output image with cell detection and segmentation is produced.

Pyramid Network (FPN) [11] along with ResNeSt-200 [17]. FPN consists of a bottom-up pathway and a top-down pathway. The bottom-up pathway extracts feature maps from the input image at different scales using a series of convolutional layers. ResNeSt-200 with deformable convolution is used as a feed-forward CNN architecture in the bottom-up pathway of the proposed approach. The top-down pathway merges feature from the bottom-up pathway using lateral connections and upsampling with features from higher-resolution layers to create a feature pyramid.

3.2 Region Proposal Network

Multi-scale features from the FPN are further processed by the Region Proposal Network (RPN), which detects the regions that contain cells and match them to the groundtruth. The matching is done by generating anchors on the input image. After the generation of anchor boxes, the next step is to associate the groundtruth bounding boxes with the generated anchors. The anchors generated are then matched to the groundtruth by taking Intersection over Union (IoU) between anchors and groundtruth. If IoU is larger than the defined threshold of 0.7, the anchor is linked to one of the groundtruth boxes and assigned to the foreground. If the IoU is greater than 0.3 and smaller than 0.7, it is considered background and otherwise ignored.

3.3 Prediction Head

At the prediction head, we have groundtruth boxes, proposal boxes from RPN, and feature maps from FPN. The job of the prediction head is to predict the class, bounding box, and binary mask for each region of interest. We are using a 3-stage Cascade Mask R-CNN [1] as the prediction head, which is an extension of Mask R-CNN [6] with the addition of cascade stages to further improve the segmentation performance. Cascade Mask R-CNN addresses the problem of making predictions that are more accurate on a pixel level. The architectures like Mask R-CNN usually malfunction while accurately detecting objects of variable quality and size in an image. This is mainly because the models are trained using a single IoU threshold i.e., 0.5, meaning that the prediction which has

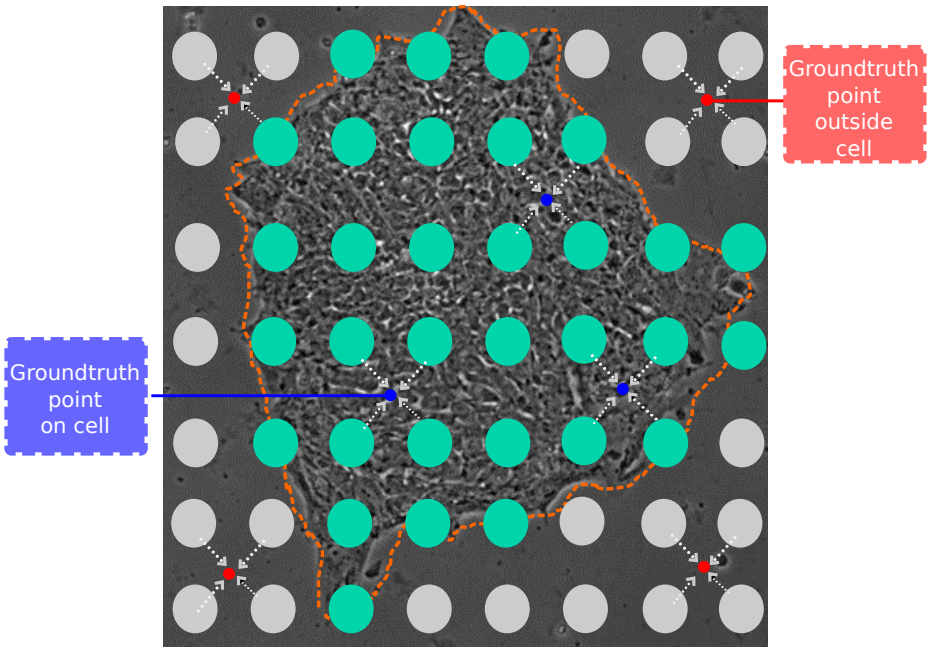


Fig. 4: **PACE weak segmentation supervision illustration.** For a 7x7 prediction mask on the regular grid (green color indicates foreground cell prediction), the predictions are obtained at the exact location of the groundtruth points with bilinear interpolation. Blue and red points indicate cell and background groundtruth points, respectively. The cell contour line is for illustration only.

over 50% match with the groundtruth will be regarded as positive samples. This can cause the model to create inaccurate proposals. To address this problem, Cascade Mask R-CNN presents a multi-stage network with the IoU threshold increasing for each stage i.e., 0.5, 0.6, and 0.7 to refine the predictions. A mask branch is added in the final stage parallel to the box branch, which is composed of a small Fully Convolutional Network (FCN) to predict a segmentation mask for each RoI in a pixel-to-pixel manner to achieve the task of instance segmentation.

In fully-supervised training, the full mask for each cell is available as the groundtruth; whereas in the proposed approach only the point labels are available as the groundtruth for training. The fully supervised method is trained by extracting a matching regular grid of labels from the groundtruth full mask. In contrast, the proposed approach uses point supervision instead of mask supervision. Predictions are approximated in the locations of the groundtruth points from the predictions on the grid using bilinear interpolation (see Fig. 4)[3]. When the prediction and the groundtruth labels are on the same point, the loss can be calculated similarly to the full supervision.

4 Dataset

In the cell biology domain, there exist numerous publically available datasets to facilitate cellular research. Among these datasets, LIVECell [4] dataset has been chosen for this study due to its size and quality. The LIVECell dataset is among the largest and most comprehensive datasets in cell biology research, comprising more than 1.6 million cells in 5,239 images. LIVECell dataset consists of eight morphologically distinct cell cultures, which makes it diverse and challenging. The average cell density in the LIVECell dataset is also very high i.e., 313 cells per image, which is almost 55 times more than the EVICAN [14] dataset.
For the training of the proposed pipeline, the full masks are discarded and replaced with different point labels. In order to analyze the impact of different point labels on the segmentation performance, six different point labels (1, 2, 4, 6, 8, 10) are generated automatically and randomly for each cell of the training data. The point can either be on the cell ('1') or anywhere inside or on the edge of the bounding box ('0').

5 Evaluation Metrics

Standard COCO evaluation protocol [12] is adapted to evaluate the performance of the proposed weakly supervised method with the same modification of the area ranges and the maximum number of detections as reported in [4]. For the evaluation, the mean average precision for both object detection and segmentation tasks at different IoU thresholds of 0.5 (mAP50), 0.75 (mAP75), and 0.5:0.95 in the steps of 0.05 (mAP) is reported. To identify the performance of the model on objects of varied sizes, we have also included mAP for different area ranges.

6 Experimental Setup

The performance of the proposed weakly supervised approach, PACE, and the state-of-the-art (SotA) method, Point2Mask [10], have been reported along with their fully supervised counterparts using the LIVECell dataset. For point-supervised weak cell segmentation, six different training experiments are reported for PACE and Point2Mask with 1-,2-,4-,6-,8-, and 10-point labels.
Training for both methods use the same settings with a learning rate of 0.02, and a momentum of 0.9 using a stochastic gradient descent-based solver. A 3x training schedule is used for the training of both methods. Anchor sizes and aspect ratios were set to 8, 16, 32, 64, 128, and 0.5, 1, 2, 3, 4 for all the settings. For data augmentation, images are flipped horizontally on a random basis to reduce the risk of over-fitting. All training used multi-scale data augmentation, meaning that image sizes were randomly changed from the original 520×704 pixels to size with the same ratios, but the shortest side was set to one of (440, 480, 520, 580, 620) pixels.
The checkpoints selection for each training was based on the validation average precision, with 4,000 being chosen for 1-, 6-, and 10-point training, 4,500 for 2-point, 3,000 for 4-point, and 5,000 for 8-point training.

Table 1: **Detection and segmentation average precision scores** on different IoU thresholds and area range for full mask supervision and $\mathcal{N}$-point supervision using different weak supervision methods i.e., **Point2Mask[10]** and the proposed method, **PACE**. The best and the second-best results for each method are represented in green and blue color, respectively.

Method	Supervision	AP		AP50		AP75		APs		APm		APl	
		Det.	Seg.	Det.	Seg.	Det.	Seg.	Det.	Seg.	Det.	Seg.	Det.	Seg.
Point2Mask [10]	Full mask	43.12	43.90	78.94	78.07	43.26	45.75	44.31	42.30	43.01	43.33	47.01	51.92
	1-point	42.67	42.37	78.71	77.58	42.46	42.96	43.91	41.33	42.16	41.37	46.19	48.64
	2-points	42.75	42.86	78.49	77.62	42.81	43.79	43.95	41.53	42.81	42.30	46.61	50.38
	4-points	**43.01**	43.17	**79.50**	77.91	42.96	44.60	43.97	41.68	**43.07**	42.77	**47.24**	51.40
	6-points	**43.32**	**43.53**	**79.69**	**78.18**	**43.31**	**44.93**	**44.54**	**42.06**	**43.31**	**43.31**	46.97	**51.52**
	8-points	42.97	**43.41**	78.86	**78.00**	**43.18**	**44.83**	43.95	**41.83**	42.54	42.77	46.94	51.44
	10-points	42.93	43.40	78.71	77.97	43.10	44.81	**44.12**	41.80	42.81	**43.04**	**47.01**	**51.65**
PACE	Full mask	48.43	47.89	81.44	80.80	51.41	51.64	48.50	45.75	49.49	48.33	54.18	56.94
	1-point	**48.87**	47.54	81.55	80.71	52.11	51.07	**48.73**	**45.48**	48.98	48.00	53.47	55.72
	2-points	48.54	47.45	**81.65**	81.03	51.67	50.87	**48.66**	**45.55**	**48.74**	47.88	53.26	55.54
	4-points	**48.56**	47.73	81.58	**80.89**	**51.83**	51.21	48.26	45.28	**49.86**	**48.60**	**54.75**	**57.33**
	6-points	47.81	47.24	81.21	80.60	50.33	50.65	47.47	44.50	48.70	48.18	**54.48**	**56.86**
	8-points	48.51	**47.81**	**81.69**	**80.88**	**51.86**	**51.74**	48.40	45.47	48.66	**48.39**	54.02	56.68
	10-points	48.18	**47.68**	81.26	80.76	51.27	**51.53**	48.12	45.40	48.68	48.22	53.59	56.77

Results Table 1 shows the overall detection and segmentation average precision scores for Point2Mask[10] and PACE on the LIVECell dataset. For the full mask supervision setting, segmentation AP scores of 43.90% and 47.89% for Point2Mask and PACE, respectively. In the case of 1-point supervision, the proposed approach, PACE, achieves an improvement of over 5% compared to Point2Mask. The performance of PACE also surpasses Point2Mask for higher levels of supervision. Specifically, for 2-, 4-, 6-, 8-, and 10-points, PACE outperforms Point2Mask by margins of 4.5%, 4.6%, 3.7%, 4.4%, and 4.3%, respectively.

7 Analysis and Discussion

In this section, we present and discuss the results of our proposed point-supervised pipeline for cell segmentation and compare it with the state-of-the-art method, Point2Mask[10]. Results in Table 1 and Fig. 5 suggest that for 6 different point labels used for training, we have achieved 98.6% to 99.8% of the fully supervised performance. Even with just 1-point label per cell instance ($\mathcal{P}_1$), we were able to achieve 99.3% of the fully supervised performance, which shows that by saving almost 83% of the time spent on full mask annotations, we can still achieve the segmentation result close to the fully supervised training. For 4-, and 10-point labels, 99.7% and 99.6% of the fully supervised performance are achieved. The best performance is observed for the 8-point label with a segmentation mAP score of 47.81%, which is 99.8% of the fully supervised performance.
In comparison to the SotA i.e., Point2Mask, PACE outperforms the best performing 6-point supervision ($\mathcal{P}_6$) with just 1-point supervision ($\mathcal{P}_1$) by a margin of 4%. Point2Mask achieved 99.16% of the fully supervised performance with 6-point labels, whereas, PACE achieves 99.83% of the LIVECell Anchor-based

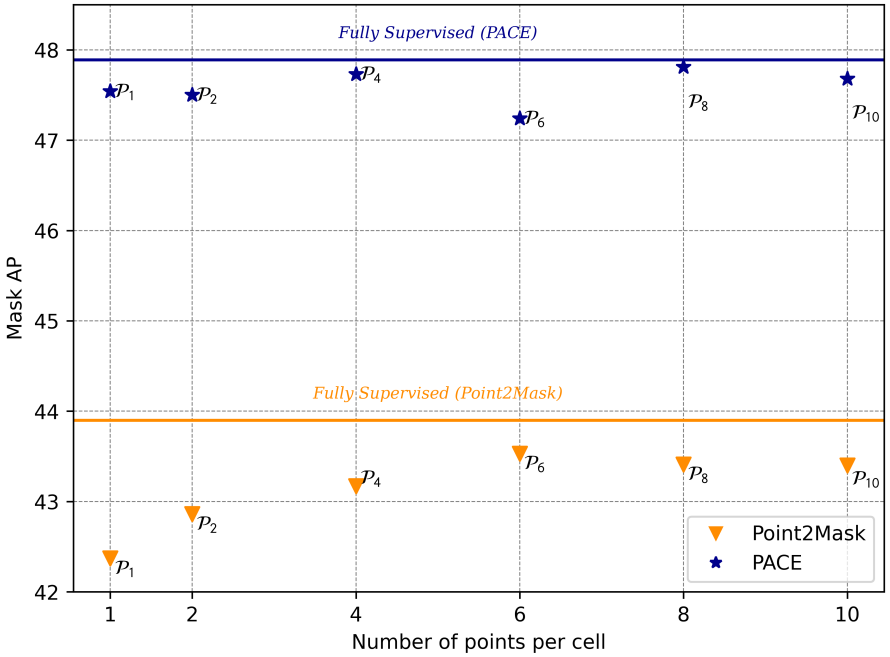


Fig. 5: **Training with different numbers of points and full mask for Point2Mask[10] and PACE.** Point2Mask results are shown in orange triangles and PACE results are shown in blue stars. PACE trained on LIVECell with as few as 1 labeled point per cell instance ($\mathcal{P}_1$) outperforms the best result of Point2Mask trained with 6 points ($\mathcal{P}_1$) by a margin of 4%. The best performance of PACE is seen for ($\mathcal{P}_8$) with Mask AP score of 47.81%.

[4] fully supervised method with 8-point labels. It shows that with only a 0.17% loss in performance, we can save almost 70% of the time spent on full mask annotations. With just a 1-point label ($\mathcal{P}_1$), Point2Mask achieved 96.51% of the fully supervised performance, whereas, the proposed method achieves 99.27%. Fig. 6 displays the comparison results of inference using the Point2Mask and PACE models on test images. These models were trained using different numbers of point annotations. Point2Mask results are depicted in the left column, while the PACE results are represented in the right column. The solid yellow lines indicate the groundtruth mask for each cell, while the dotted red lines represent the predictions made by the model. The rows colored in red, gray, and blue represent the inference results obtained from models trained with 2, 6, and 8 points, respectively. The label "AP50" displayed on top of each prediction sub-image denotes the segmentation average precision score at the IoU threshold of 0.5. The top row represents the results for the models trained with 2-point labels. The inference result of the Point2Mask model reveals some instances of

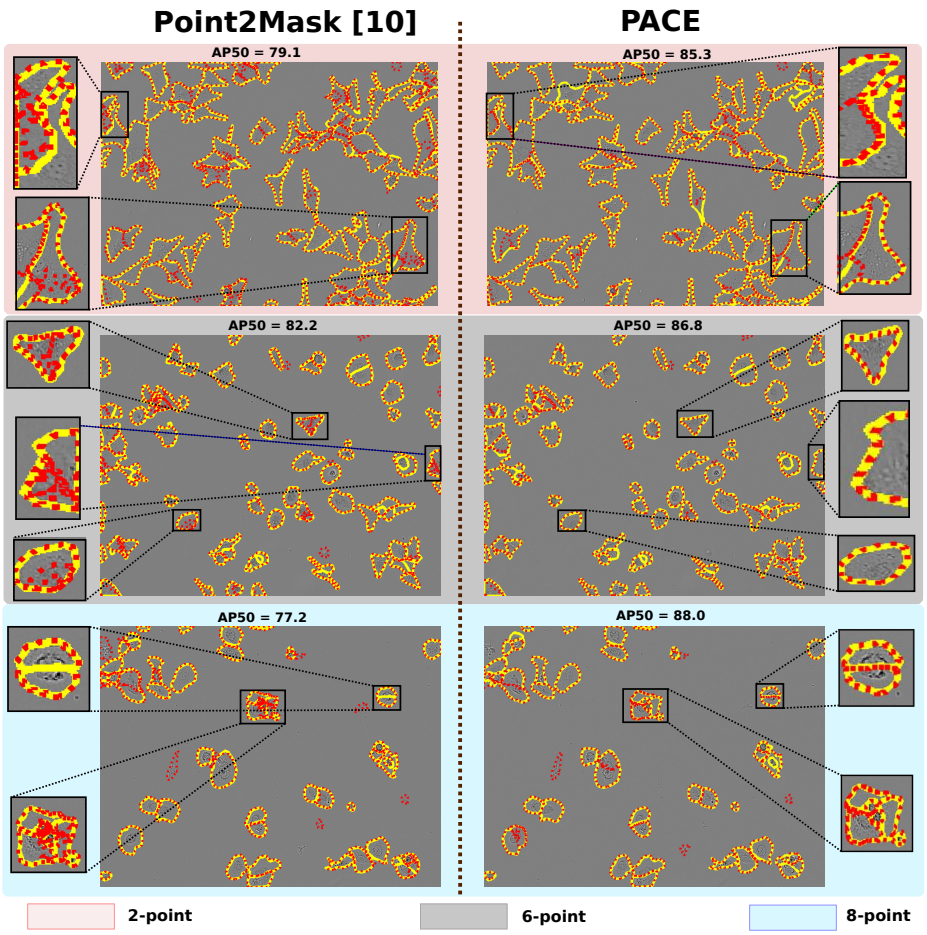


Fig. 6: **Inference results** using different point labels for **Point2Mask** [**10**] and the proposed approach, **PACE**. The Point2Mask and PACE results are shown in the left and right columns, respectively. Groundtruth masks are represented by solid yellow lines while dotted red lines show the predictions made by the models. The red, gray, and blue rows represent the inference results obtained from the models trained on 2, 6, and 8-point annotations, respectively. Some cell instances where Point2Mask failed and PACE achieved better segmentation results are also highlighted.

false positives and splitting of a large cell into two. In contrast, the proposed approach, PACE, exhibits better performance in such cases. To provide a clearer picture of the results, specific parts of the images where the SotA Point2Mask model failed to segment cells accurately were zoomed in. It can be observed that PACE is able to segment these specific parts of the images with more precision. The middle row (colored in gray) shows the results obtained from the models

trained with 6-point labels. It is evident that the proposed weak cell supervision approach performs relatively well with an AP50 score of 86.8. The last row depicts the inference results for the model trained with 8-point labels. PACE outperforms Point2Mask, as the latter segments two cells as one in one instance, while PACE correctly identifies them as two separate cells. Both methods exhibit some false positives, and in a few cases, the groundtruth is unavailable for certain cells.

The proposed approach has enabled us to achieve performance that is close to full supervision while significantly reducing the time required to annotate the data compared to full mask annotation. Results indicate that even with 1-point supervision during training, we can achieve over 99.3% of the performance achieved with full supervision. Moreover, the proposed approach can also reduce the level of expertise required from biologists to establish cell boundaries. By reducing the time and effort required for data annotation, the proposed method allows for the analysis of more data in a shorter amount of time. The proposed approach can be scaled up to larger and more complex datasets without a corresponding increase in the amount of manual labor and expertise required for data annotation. This increased efficiency in data analysis could lead to a better understanding of biological and medical phenomena, potentially leading to the development of new treatments and diagnostic tools.

8 Conclusion

PACE provides an improved approach for weakly supervised cell segmentation using point labels for training instead of the full mask. With just a 1-point label, more than 80% of the time spent on full mask annotations can be saved with just a 0.7% loss in performance compared to the fully supervised method. By utilizing the results of this study, we have demonstrated that it is possible to decrease the time and costs associated with fully annotating the data. In addition, we can also minimize the level of expert knowledge required from biologists to establish cell boundaries. The proposed point-supervised approach can also potentially increase the scalability of cell segmentation studies in biology and medicine. Using the proposed approach, a substantial amount of unlabeled image-based cellular data can be utilized, which in turn can help in conducting larger-scale studies and analyses. The proposed approach could further the research in the field of biology and medicine, potentially leading to new discoveries, as more data can be analyzed in a shorter amount of time.

References

1. Cai, Z., Vasconcelos, N.: Cascade r-cnn: Delving into high quality object detection. In: Proceedings of the IEEE conference on computer vision and pattern recognition (2018)
2. Chen, Z., Chen, Z., Liu, J., Zheng, Q., Zhu, Y., Zuo, Y., Wang, Z., Guan, X., Wang, Y., Li, Y.: Weakly supervised histopathology image segmentation with sparse point annotations. IEEE Journal of Biomedical and Health Informatics (2020)

3. Cheng, B., Parkhi, O., Kirillov, A.: Pointly-supervised instance segmentation. arXiv preprint arXiv:2104.06404 (2021)
4. Edlund, C., Jackson, T.R., Khalid, N., Bevan, N., Dale, T., Dengel, A., Ahmed, S., Trygg, J., Sjögren, R.: Livecell—a large-scale dataset for label-free live cell segmentation. Nature methods (2021)
5. Guerrero-Peña, F.A., Fernandez, P.D.M., Ren, T.I., Cunha, A.: A weakly supervised method for instance segmentation of biological cells. In: Domain Adaptation and Representation Transfer and Medical Image Learning with Less Labels and Imperfect Data. Springer (2019)
6. He, K., Gkioxari, G., Dollár, P., Girshick, R.: Mask r-cnn. In: Proceedings of the IEEE international conference on computer vision (2017)
7. Khalid, N., Koochali, M., Rajashekar, V., Munir, M., Edlund, C., Jackson, T.R., Trygg, J., Sjögren, R., Dengel, A., Ahmed, S.: Deepmucs: A framework for co-culture microscopic image analysis: From generation to segmentation. In: 2022 IEEE-EMBS International Conference on Biomedical and Health Informatics (BHI). IEEE (2022)
8. Khalid, N., Munir, M., Edlund, C., Jackson, T.R., Trygg, J., Sjögren, R., Dengel, A., Ahmed, S.: Deepcens: An end-to-end pipeline for cell and nucleus segmentation in microscopic images. In: 2021 International Joint Conference on Neural Networks (IJCNN). IEEE (2021)
9. Khalid, N., Munir, M., Edlund, C., Jackson, T.R., Trygg, J., Sjögren, R., Dengel, A., Ahmed, S.: Deepcis: An end-to-end pipeline for cell-type aware instance segmentation in microscopic images. In: 2021 IEEE EMBS International Conference on Biomedical and Health Informatics (BHI). IEEE (2021)
10. Khalid, N., Schmeisser, F., Koochali, M., Munir, M., Edlund, C., Jackson, T.R., Trygg, J., Sjögren, R., Dengel, A., Ahmed, S.: Point2mask: A weakly supervised approach for cell segmentation using point annotation. In: Medical Image Understanding and Analysis: 26th Annual Conference, MIUA 2022, Cambridge, UK, July 27–29, 2022, Proceedings. Springer (2022)
11. Lin, T.Y., Dollár, P., Girshick, R., He, K., Hariharan, B., Belongie, S.: Feature pyramid networks for object detection. In: Proceedings of the IEEE conference on computer vision and pattern recognition (2017)
12. Lin, T.Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., Zitnick, C.L.: Microsoft coco: Common objects in context. In: European conference on computer vision. Springer (2014)
13. Papadopoulos, D.P., Uijlings, J.R., Keller, F., Ferrari, V.: Extreme clicking for efficient object annotation. In: Proceedings of the IEEE international conference on computer vision (2017)
14. Schwendy, M., Unger, R.E., Parekh, S.H.: Evican–a balanced dataset for algorithm development in cell and nucleus segmentation. Bioinformatics (2020)
15. Stringer, C., Wang, T., Michaelos, M., Pachitariu, M.: Cellpose: a generalist algorithm for cellular segmentation. Nature Methods (2020)
16. Tsai, H.F., Gajda, J., Sloan, T.F., Rares, A., Shen, A.Q.: Usiigaci: Instance-aware cell tracking in stain-free phase contrast microscopy enabled by machine learning. SoftwareX (2019)
17. Zhang, H., Wu, C., Zhang, Z., Zhu, Y., Lin, H., Zhang, Z., Sun, Y., He, T., Mueller, J., Manmatha, R., et al.: Resnest: Split-attention networks. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (2022)
18. Zhou, Y., Zhu, Y., Ye, Q., Qiu, Q., Jiao, J.: Weakly supervised instance segmentation using class peak response. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (2018)