Cell Instance Segmentation with Large Vision Model for Circulating Aberrant Cells Identification through Fluorescence In Situ Hybridization Images

Anonymous Author(s)
Affiliation
Address
email

Abstract

Liquid biopsy presents a promising non-invasive or minimally invasive approach for early detection of lung cancer. By quantifying 4-color fluorescence in situ hybridization (FISH) signals, circulating genetically abnormal cells (CACs) can potentially be identified with high stability, sensitivity, and specificity. However, precise segmentation of cells is a prerequisite for accurate signal counting. In supervised learning, deep learning models have shown excellent instance segmentation capabilities but require substantial labeled data. To overcome this limitation, we propose a novel zero-shot learning framework for cell instance segmentation. Specifically, we first leverage the Watershed algorithm to generate segmentation proposals and prompts for a large generative vision model, the Segment Anything Model (SAM). We then filter the model outputs based on prior knowledge to obtain pseudo labels for training a specialized instance segmentation model. Our approach eliminates the need for manually labeled data. We demonstrate its effectiveness by segmenting cell images from liquid biopsy and comparing performance against generalized cell segmentation methods(Cellpose). This zero-shot learning paradigm could expand the applicability of vision models to specialized medical imaging applications without costly labeling.

1 Introduction

Lung cancer, the second most commonly diagnosed malignancy globally, poses a significant clinical challenge due to this often-late diagnosis. Liquid biopsy[11], a minimally invasive or non-invasive technology, offers a promising approach for early detection by analyzing circulating biomarkers in body fluids. One such approach involves identifying circulating genetically aberrant cells(CAC) through fluorescence in situ hybridization(FISH), a method that detects chromosomal abnormalities. According to Katz's[6] criteria, the presence of three or more CACs per 100ul of blood indicates an increased risk for stage I or II lung cancer.

Accurate segmentation of cells in liquid biopsy images is crucial for identifying CACs. However, this process is labor-intensive due to the high number of cells often present within a signal field of view. Manual labeling for training segmentation models is time-consuming and resource-intensive. Therefore, developing methods that significantly reduce human labeling efforts is critical for advancing liquid biopsy applications in lung cancer diagnosis.

Cellpose is a generalized, deep learning-based cell segmentation method developed by MouseLand. Unlike watershed which uses grayscale to compute gradients, image gradients in Cell pose are created by simulating diffusion from GT, using a thermal diffusion model for each ROI, and iterative diffusion from the center of the ROI to sub-simulate to create a topology map. MouseLand not only provides a

Submitted to 38th Conference on Neural Information Processing Systems (NeurIPS 2024). Do not distribute.

variety of base models for different segmentation scenarios but also supports convenient fine-tuning methods in Cellpose version 2.0.

2 Related Work

2.1 Machine Vision Algorithm for Cell Segmentation

The field of cell instance segmentation has traditionally relied on region-based image segmentation techniques based on direct region finding. These methods partition the image into distinct regions based on various image features and subsequently classify each region as belonging to a specific object or background. Three prominent approaches include:

Region Growing [1] This algorithm iteratively incorporates neighboring pixels around a seed pixel with a high-intensity value that meets a predetermined similarity criterion, ultimately forming a complete region representing a cell. While its simplicity and low computational cost make it attractive, complex image backgrounds can higher its accuracy.

Region Splitting and Merging [2] This approach starts with the entire image and iteratively subdivides it into smaller regions. Subsequently, adjacent regions with similar features are merged, ultimately resulting in the desired segmentation masks. This method offers flexibility in handling diverse image structures but can be computationally expensive for large datasets.

Watershed Algorithm [3] Inspired by the behavior of water flowing downhill, this method treats the image as a topographical landscape, with pixel intensities representing elevation. Local minima(valleys) act as watershed, and the algorithm segments the image into distinct regions based on the flow paths towards these minima. While effective for certain scenarios, it might struggle with noisy images or intricate cell shapes.

2.2 Large Vision Model(LVM) for instance segmentation

Recent years have witnessed a significant shift in the field of computer vision, driven by the remarkable success of deep learning models, particularly LVMs. These models leverage the power of deep neural networks and massive datasets to achieve unprecedented levels of accuracy and generalization in various tasks, including instance segmentation.

One such LVM, the Segment Anything Model(SAM)[7] utilizes separate encoders for images and prompts to generate a shared representation space. This representation is then fed into a lightweight mask decoder for predicting segmentation masks. SAM demonstrates effective segmentation capabilities, paving the way for further advancements in instance segmentation.

Building upon SAM's foundation, subsequent models like SEEM[14] address its limitations by incorporating a border range of interactional and semantic information, leading to even more robust and comprehensive segmentation performance. Additionally, MedSAM[10] extends SAM for the specific domain of medical image segmentation, showcasing its potential in real-world applications. AutoSAM, which generates a fully automated solution for SAM's cues, creates alternative cues based on a network of auxiliary cueing encoders for the input image having SAM. MedSAM proposes an annotation process for medical datasets using SAM and introduces a small number of localization frameworks. MedLSAM[8] significantly reduces the annotation burden by automatically identifying the target anatomical regions throughout the dataset to be annotated. SAMAug[4] uses a new Visual Point augmentation framework for generating additional point prompts without the need of additional manual operations to the SAM to perform additional manual operations.

2.3 Instance Segmentation Networks and Evaluation Metrics

Instance segmentation aims to not only identify the presence of cells but also delineate their individual boundaries, making it a more intricate task compared to traditional segmentation. A widely adopted model, for instance, segmentation is the Mask R-CNN[5], which combines the strengths of both object detection and semantic segmentation. Its ability to accurately segment complex cell structures has established it as a prominent tool in this field. Modified the cell segmentation field is well

addressed by mask R-CNN. Fixed-size region proposals and attention techniques were employed by MACD R-CNN[9] to find aberrant nuclei.

Cellpose[13] is a generalized, deep learning-based cell segmentation method developed by MouseLand. Unlike watershed which uses grayscale to compute gradients, image gradients in Cell pose are created by simulating diffusion from GT, using a thermal diffusion model for each ROI, and iterative diffusion from the center of the ROI to sub-simulate to create a topology map. MouseLand not only provides a variety of base models for different segmentation scenarios but also supports convenient fine-tuning methods in Cellpose version 2.0[12].

Another crucial aspect of evaluating the performance of segmentation models is the use of appropriate similarity metrics, Two widely employed metrics are the Dice coefficient and Intersection over Union(IoU).

Dice coefficient This metric focuses on the overlap between predicted and ground truth segmentation masks, offering a pixel-wise comparison. It calculates the ratio of twice the area of the intersection between predicted and ground truth masks to the sum of their areas. A higher Dice coefficient indicates better segmentation accuracy, particularly for smaller objects like cells.

IoU This metric measures the ratio of the area of overlap between predicted and ground truth masks to the area of their combined union.

The dice coefficient was defined as:

$$Dice = \frac{2|\mathrm{Pred} \cap \mathrm{GT}|}{|\mathrm{Pred}| + |\mathrm{GT}|} \tag{1}$$

IoU was defined as:

$$IoU = \frac{|\mathrm{Pred} \cap \mathrm{GT}|}{|\mathrm{Pred} \cup \mathrm{GT}|} \tag{2}$$

Where Pred denotes the model that predicts the region and GT represents the ground truth.

Given the typically small size and intricate shapes of cells, the Dice coefficient is the preferred evaluation metric in this specific task. Its sensitivity towards accurately capturing the complete cell region makes it a more reliable indicator of performance compared to the IoU metric, which might be less sensitive to small deviations in segmentation boundaries.

3 Method

3.1 Data acquisition and pre-processing

Five channels, each with 4128*3600 pixels, were used to capture the data for each group using a 4-color fluorescence in situ hybridization(FISH) probe set from Sanmed Biotechnology Inc. in Zhuhai, China. These channels include DAPI, aqua, green, gold, and red. FISH is a cytogenetic technique used to visualize specific DNA sequences in situ. The 4-color FISH probe set used in this study targets specific regions of interest within the cell nuclei, allowing for their visualization and subsequent segmentation. In this task, we only use the DAPI channel for segmentation as it specifically stains the cell nuclei, providing a clear contrast between the nuclei and the background.

To improve image quality and remove background noise, we employed Otsu's thresholding technique. This method automatically determines an optimal threshold to separate the foreground(cell nuclei) from the background based on maximizing the between-class variance. Otsu's thresholding technique has been demonstrated to be effective in segmenting cell nuclei in various image analysis tasks.

3.2 Prompt generate

As shown in Figure 1, The pipeline of our framework. Prompt generation flow is shown in the red region, the pseudo-label filter is shown in the green region, and the red region represents the prompt generation stage. This process involves two steps to obtain the complete prompt:

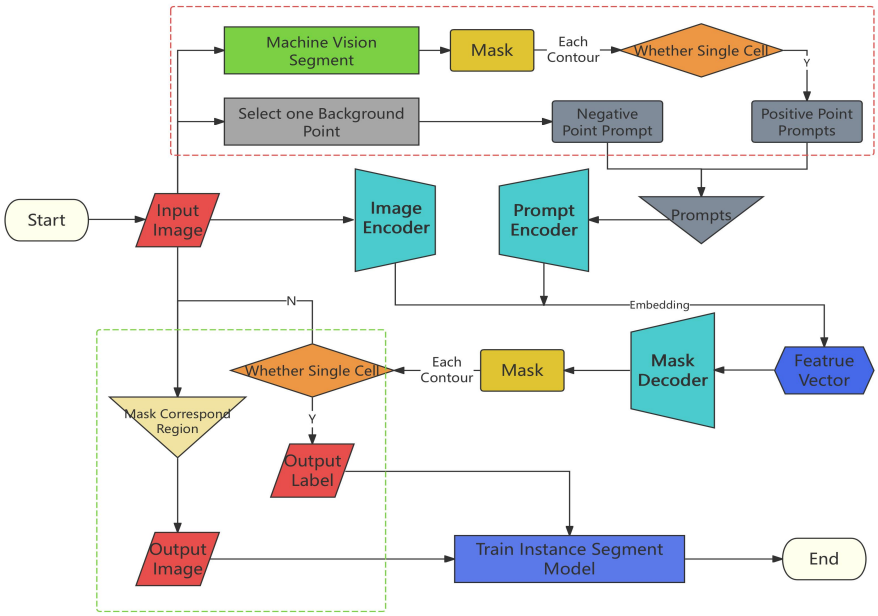

Figure 1: The pipeline of our zero-shot cell instance segmentation framework using a large vision model. The process consists of three main stages: 1) Prompt generation (red region): Machine vision algorithms generate coarse cell segmentation, providing positive point prompts for individual cells. A negative point prompt is selected from the background. 2) Segmentation: The input image and generated prompts are encoded and processed by the Segment Anything Model (SAM) to produce instance segmentation masks. 3) Pseudo-label filtering (green region): Each segmented contour is analyzed to determine if it represents a single cell. Valid single-cell masks are used as pseudo-labels, while invalid masks are excluded. The resulting pseudo-labeled dataset is used to train a specialized instance segmentation model (e.g., Mask R-CNN) for efficient and accurate cell segmentation.

Coarse cell segmentation. A machine vision algorithm, specifically region growing, region splitting and merging, and watershed is used to roughly segment the cell regions. Since cells are morphologically similar in size, individual cells can be effectively filtered by calculating features like area, length, width, and roundness of the contours. The center points of the segmented individual cell contours are then saved as positive point prompts.

Negative point prompt. According to studies on prompts segmentation properties of the Segment Anything Model(SAM), at least one negative point prompt pointing to the background is necessary for better cell contour segmentation Therefore the point with the smallest gray value in the image is selected as the negative point prompt. This negative point helps the model distinguish between foreground(cell) and background region during segmentation.

3.3 Segment and Pseudo Label Filtering

Segmentation is performed using the SAM based on the prompts generated in the previous step. SAM's ability to encode both positive and negative point prompts, along with the powerful image segmentation capabilities of the underlying large vision model, allows it to effectively segment individual cells, addressing the limitations of traditional vision algorithms that often struggle with aggregated cells.

After SAM segmentation, each contour is further analyzed to determine if it represents a single cell. If the features confirm it is a single cell, it is added to the pseudo-labeled set. However, if the contour likely represents aggregated cells and cannot be effectively segmented, it is excluded from the pseudo-labeled set, and the corresponding region in the original image is filled with background pixels. This filtering process helps ensure the accuracy of the pseudo-labeled data.

Method	Dice(%)	Times per frames(s)
SAM	52.48	37,33
SAM(Prompt By Watershed)	60.03	10.93
Cellpose(Default)	74.71	21.64
Cellpose(Fine-tune)	80.22	27,43
Mask R-CNN(Label By Our Method)	87.65	4.19
Mask R-CNN(Label By Human)	93.51	4.19

Table 1: Evaluation metrics of different methods for segmentation

Through this approach, we obtain a batch of high-quality pseudo-labels without any manual annotation, significantly reducing workload and accelerating the training process. These pseudo-annotations are then used to train a small instance segmentation model, further reducing the need for manual annotations.

4 Results

In our experiment, we employed the SAM-B model as a generalized segmentation large model, the Watershed algorithm for traditional machine vision segmentation, and Mask R-CNN for instance segmentation. For a baseline, we fine-tune a pre-existing nuclei model from MouseLand on our training set.

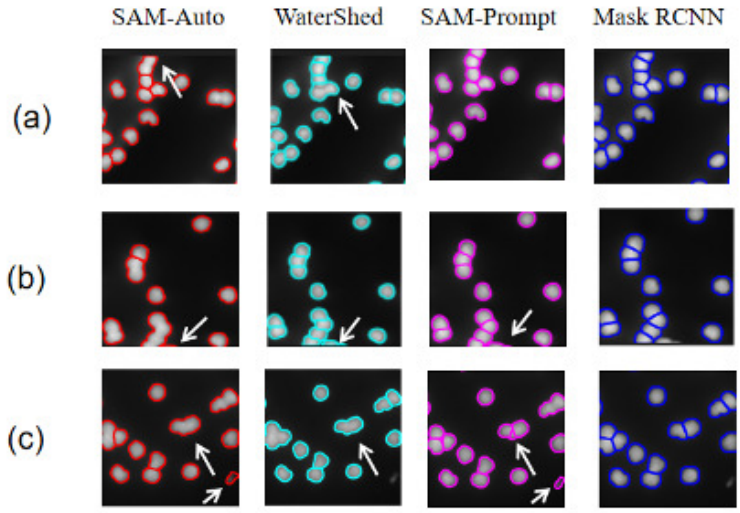

Figure 2: Comparison for the typical cell instance segmentation performance by different algorithms. (a): SAM-Auto and Watershed exhibit lower segmentation accuracy compared to SAM with Prompt.(b): All untrained methods(SAM-Auto, Watershed, and SAM-Prompt) struggle to segment the noisy object effectively. (c): SAM with Prompt exhibits over-segmentation.

Table 1 demonstrates that the Watershed algorithm outperforms SAM's automatic segmentation in this specific task. This aligns with expectations; while zero-shot models excel in broader generalization, traditional algorithms can be more effective in constrained scenarios. However, when provided with prompts, SAM exhibits superior segmentation, highlighting the knowledge-driven strength of zero-shot approaches.(See Fig.22)

The Watershed algorithm operates on grayscale values, making it sensitive to image quality and noise. Conversely, Cellpose, as a dedicated cell segmentation model, leverages extensive pre-training on diverse cellular data. This pre-training grants it an edge over the Watershed algorithm with default parameters, emphasizing the power of large-scale model training for specialized tasks. Notably,

fine-tuning Cellpose further improved performance, illustrating the value of domain adaptation in zero-shot scenarios.

Finally, our framework's Mask R-CNN model demonstrated both superior segmentation performance and the shortest computational time. This finding has significant implications for real-world biological applications where rapid and accurate cell segmentation is crucial, as it reduces annotation burdens and supports time-sensitive analysis.

5 Conclusion

In conclusion, this work explores the application of zero-shot learning with a large vision model(LVM) for cell instance segmentation. Our finding highlights the potential of this method to address the challenges of scarce labeled data and intensive computational demands associated with supervised training methods. By leveraging a pre-trained, lightweight model informed by prior knowledge of cell morphology, we achieved accurate segmentation without human labeling intervention. This framework paves the way for efficient and scalable cell segmentation, not only in biology but also in other domains where prior knowledge can guide object identification. Future research should focus on exploring diverse LVM architectures and enriching knowledge representations to further enhance accuracy and generalizability across various biological contexts.

References

[1] R. Adams and L. Bishof. Seeded region growing. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1994.

[2] Kanur Aneja, Florence Laguzet, Lionel Lacassagne, and Alain Merigot. Video-rate image segmentation by means of region splitting and merging. In *IEEE International Conference on Signal amp; Image Processing Applications*, 2009.

[3] S. Beucher and Centre De Morphologie Mathmatique. The watershed transformation applied to image segmentation. *Scanning microscopy. Supplement*, 6:299–314, 2000.

[4] Haixing Dai, Chong Ma, Zhiling Yan, Zhengliang Liu, Enze Shi, Yiwei Li, Peng Shu, Xiaozheng Wei, Lin Zhao, Zihao Wu, et al. Samaug: Point prompt augmentation for segment anything model. *arXiv preprint arXiv:2307.01187*, 2023.

[5] Kaiming He, Georgia Gkioxari, Piotr Dollár, and Ross Girshick. Mask r-cnn. In *Proceedings of the IEEE international conference on computer vision*, pages 2961–2969, 2017.

[6] Ruth L. Katz, Tanweer M. Zaidi, and Deep PujaraNamita D. ShanbhagDuy TruongShekhar PatilReza J. MehranRa A. El-ZeinSanjay S. SheteJoshua D. Kuban. Identification of circulating tumor cells using 4-color fluorescence in situ hybridization: Validation of a noninvasive aid for ruling out lung cancer in patients with low-dose computed tomography-detected lung nodules. *Cancer: A Journal of the American Cancer Society*, 128(8):553–562, 2020.

[7] 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.

[8] Wenhui Lei, Xu Wei, Xiaofan Zhang, Kang Li, and Shaoting Zhang. Medlsam: Localize and segment anything model for 3d medical images. *arXiv preprint arXiv:2306.14752*, 2023.

[9] Baoyan Ma, Jian Zhang, Feng Cao, and Yongjun He. Macd r-cnn: An abnormal cell nucleus detection method. *IEEE Access*, PP(99):1–1, 2020.

[10] Jun Ma, Yuting He, Feifei Li, Lin Han, Chenyu You, and Bo Wang. Segment anything in medical images. *Nature Communications*, 15(1), 2024.

[11] Valerie Poulet, Josephine Taly. Liquid biopsy: General concepts. *Acta Cytologica: The Journal of Clinical Cytology and Cytopathology*, 63(6), 2019.

[12] C. Stringer and Marius Pachitariu. Cellpose 2.0: how to train your own model. *Nature Methods*, 19:1634 – 1641, 2022.

[13] Carsen Stringer, Tim Wang, Michalis Michaelos, and Marius Pachitariu. Cellpose: a generalist algorithm for cellular segmentation. *Nature Methods*, 18(1):1–7, 2021.

[14] Xueyan Zou, Jianwei Yang, Hao Zhang, Feng Li, Linjie Li, Jianfeng Gao, and Yong Jae Lee. Segment everything everywhere all at once. *ArXiv*, abs/2304.06718, 2023.