Visual-Recognition-using-Deep-Learning

Final Project-Report

Team: 5
Member:
313553044 江仲恩
313551139 陳冠豪
313553046 馬國維
313553037 黃瑜明

Topic:

Sartorius - Cell Instance Segmentation

Github Link:

https://github.com/CEJiang/DLCV-Final-Project

1. Introduction

Neurological diseases, such as Alzheimer's disease, are significant causes of death and disability. Traditionally, these diseases have been detected through manual observation using light microscopes, which offers the advantages of accessibility and non-invasiveness. However, segmenting individual neuronal cells in microscopic images is both challenging and time-intensive, particularly when cells exhibit complex morphological variations and overlapping distributions.

In cell segmentation tasks, traditional semantic segmentation can only identify which pixels belong to which category (such as cell or background) but cannot distinguish between different individual cells. In contrast, **instance segmentation** not only identifies cell regions but, more importantly, can distinguish each individual cell, assigning a unique identifier to each cell. This capability is crucial for cell counting, tracking, and morphological analysis.

The appearance of neuronal cells is varied and irregular. Among eight different cancer cell types tested, SH-SY5Y human neuroblastoma cells consistently exhibit the lowest precision scores, highlighting the particular difficulty in segmenting this cell type.

Computer vision technology, especially deep learning-driven instance segmentation, provides new possibilities for addressing these challenges.

In this competition, we obtained images containing multiple neuronal cells, including SH-SY5Y and two other types of neuronal cells. Our goal is to use these images to train and test models to achieve higher accuracy in neuronal cell instance segmentation, providing technical support for automated diagnosis and research of neurological diseases.

2. Related works

2.1 Cellpose

Cellpose[1] is a specialized deep learning model architecture for cell segmentation, trained on a dataset containing over 70,000 segmented objects from cellular images. The core innovation of Cellpose lies in its flow field prediction approach, which establishes a topological map to compute gradients in x and y directions, forming vector fields that point toward each cell's center. By tracking these vector fields, all pixels belonging to the same cell eventually converge to the same fixed point, enabling precise instance segmentation.

Flow Field Generation Mechanism

Cellpose uses a simulated heat diffusion process to create target vector fields:

1. Heat sources are placed at each cell's center point
2. Heat diffusion simulation is performed within cell boundaries using "leaky" boundary conditions
3. After sufficient iterations, the system reaches thermal equilibrium, forming an energy function
4. Horizontal and vertical gradients of this energy function are computed to obtain vector fields pointing toward cell centers

Network Architecture Design

Cellpose is based on a modified U-Net architecture with encoder-decoder structure and skip connections. Key innovations include:

- Residual blocks replacing standard convolutional layers
- Additive fusion instead of feature concatenation to reduce parameters
- Global style vectors for adaptation to different image types
- Simultaneous prediction of three outputs: horizontal flow field, vertical flow field, and cell probability map

Instance Recovery Process

During testing, the predicted vector fields are used to construct a dynamical system. Starting from each pixel, iterative movement along predicted gradient directions continues until convergence to fixed points. All pixels converging to the same fixed point are classified as the same cell instance. This approach better handles irregular cell shapes and overlapping boundaries compared to traditional watershed algorithms or direct mask prediction.

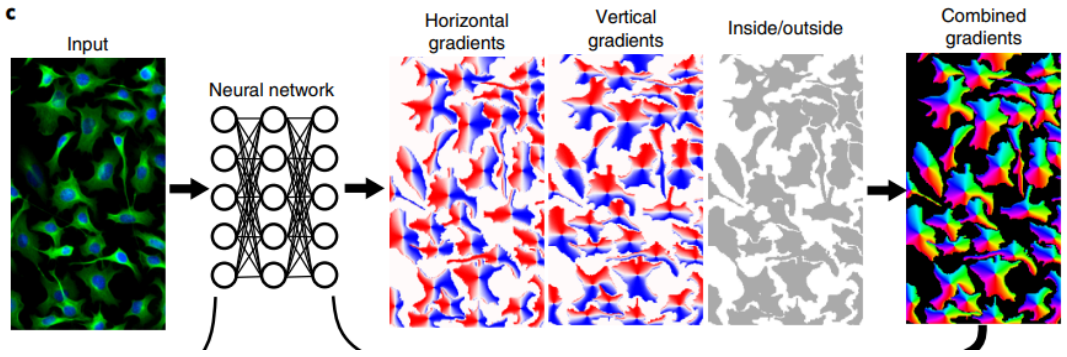


Fig.1 Cellpose Architecture[1]

2.2 Cellpose-SAM

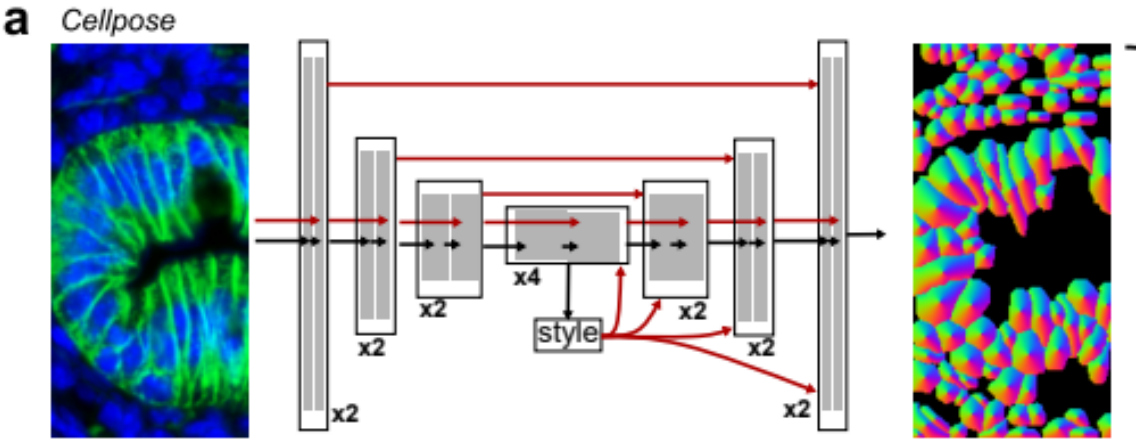
Cellpose-SAM[2] is a novel model that combines Cellpose with the Segment Anything Model (SAM), aimed at improving the generalization capabilities of cell segmentation.

Architectural Innovations

Cellpose-SAM adopts SAM's Vision Transformer (ViT-L) Encoder to replace the original Cellpose U-Net encoder. Key improvements include:

- Input size adjustment from 1024×1024 to 256×256
- Patch size modification from 16×16 to 8×8
- Reverting local attention layers to ViT-L's default global attention mechanism

These improvements enable the model to better capture global contextual information of cells while maintaining computational efficiency.



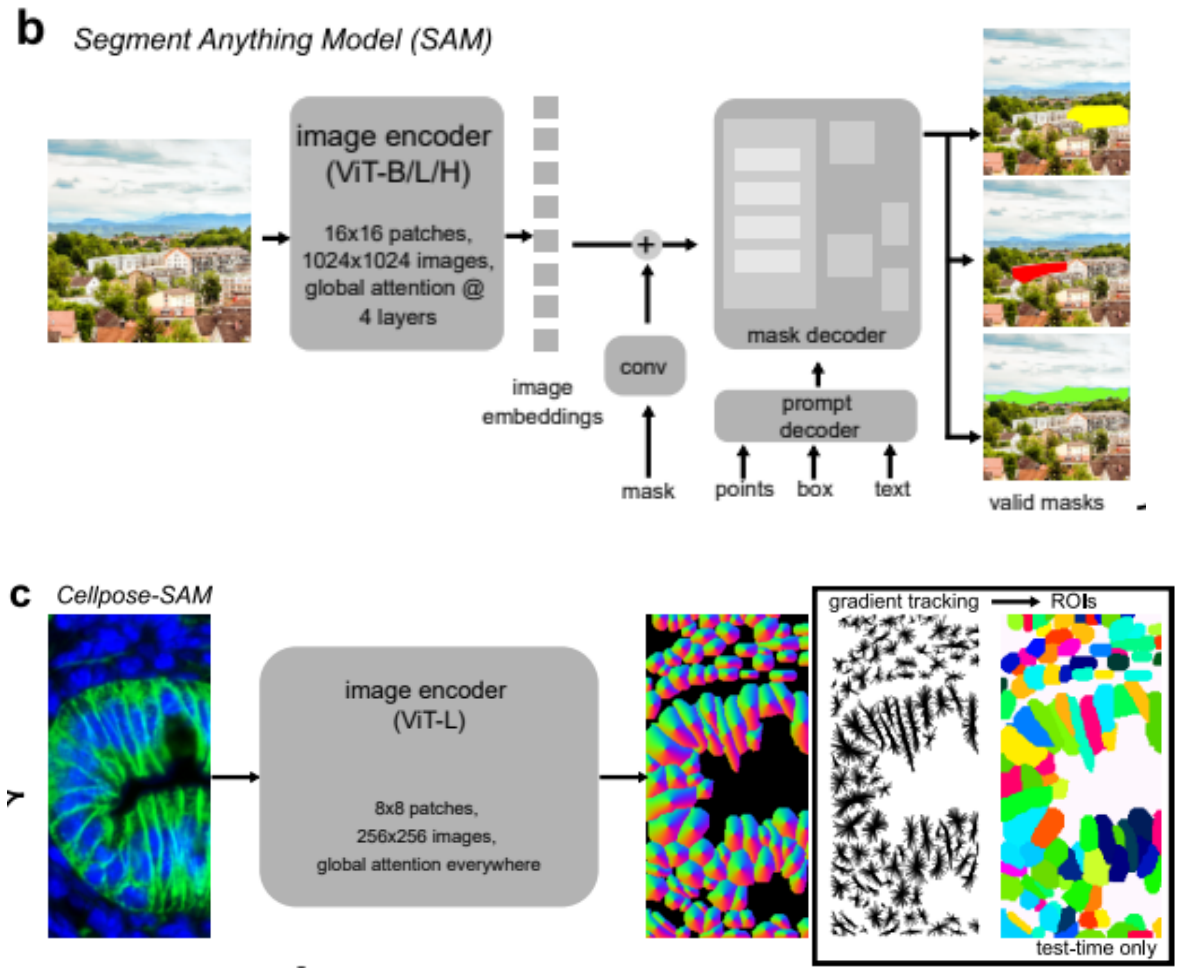


Fig.2 Cellpose-SAM Architecture[2]

2.3 Detectron2

Detectron2 is a next-generation object detection and instance segmentation framework developed by Facebook AI Research, offering enhanced performance and flexibility compared to the original Detectron[3]. The framework integrates multiple advanced instance segmentation algorithms, including Mask R-CNN, Cascade R-CNN, and PointRend, providing robust technical support for cell segmentation tasks.

Core Architectural Features

Detectron2 adopts a modular design that decomposes the entire detection pipeline into configurable components:

- **Backbone Networks**: Support for various backbone networks including ResNet, ResNeXt, and FPN
- **Region Proposal Networks (RPN)**: Efficient candidate region generation mechanism
- **ROI Heads**: Pluggable region-of-interest processing modules

- **Loss Functions**: Support for multiple loss function combinations and weight adjustments

Instance Segmentation Algorithms

Mask R-CNN: Extends Faster R-CNN by adding a mask branch, enabling simultaneous object detection and pixel-level segmentation[6]. For cell segmentation, Mask R-CNN can precisely identify cell boundaries and generate high-quality segmentation masks.

Pre-training Strategies

Detectron2 provides models pre-trained on the COCO dataset, whose learned feature representations can be effectively transferred to cell segmentation tasks. Through further pre-training on large-scale cell datasets such as LiveCell, models can better adapt to cellular morphological features and distribution patterns.

2.4 Copy-Paste

Copy-Paste[4] is an effective data augmentation method for instance segmentation. This technique copies objects from one image and pastes them into another image, effectively increasing training data diversity. In cell segmentation tasks, this method is particularly useful as it can simulate cell distribution patterns in different environments.

3. Method / Approach

3.1 Pretrain Stage

We use LiveCell_dataset_2021 as our pretrain dataset. The dataset contains 8 different types of cell including SH-SY5Y and the other type of cell.

3.2 Data Preprocessing

In this stage, we will do some additional data augmentation for cellpose model. For detectron2 model, we found that the perform will get higher if not do the augmentation. Thus, we only do a randomCrop and data type conversion.

Augmentation Method	Parameter Settings	Description
Copy-Paste	keep_prob=0.5, select_prob=0.9, occ_thresh=0.7	Copy-paste augmentation to enhance instance diversity
HorizontalFlip	p=0.5	Horizontal flipping
VerticalFlip	p=0.5	Vertical flipping
Affine	scale=(0.7, 1.3) translate_percent= {'x': (-0.15, 0.15), 'y': (-0.15, 0.15)} rotate=(-180, 180) p=0.5	Affine transformations including scaling, translation, and rotation
OneOf(ElasticTransform ,GridDistortion)	ElasticTransform(alpha=50, sigma=50 * 0.05, p=1.0), GridDistortion(num_steps=5, distort_limit=0.1, p=1.0), p=0.2	Elastic transformation or grid distortion
RandomBrightnessCont rast	brightness_limit=0.2 contrast_limit=0.2 p=0.3	Random brightness and contrast adjustment
GaussianBlur	blur_limit=(3, 7) p=0.25	Gaussian blur
ToFloat, ToTensorV2		Data type conversion

3.3 Model training

Cellpose-SAM Model

Model Architecture:

We utilize Cellpose-SAM provided from cpsam[5]. Since the input of the model is a 256x256 shape. randomly crops input images to 256×256 pixels, we designed a sliding window approach to handle full-resolution images.

Sliding Window Strategy:

- Original image size: 520×704 pixels
- Segmentation method: Split images into 4×4 patches with stride=73×134
- Each patch size: 301×302 pixels

Loss Function: The Cellpose loss function consists of two components:

- **Flow Loss**: Computes mean squared error between predicted flow and ground truth flow
- **CellProb Loss**: Applies binary cross-entropy loss between ground truth and predicted cell probability maps

Detectron2

Model Architecture:
We employ the R50-FPN (ResNet-50 with Feature Pyramid Network) backbone with lr-sched=3x learning rate schedule[3].

Multi-scale patch:

- Process images at 6 different scales (440, 480, 520, 560, 580, 620 pixels)
- Split large images into overlapping patches (301×302 pixels) with stride (73×134) to prevent out-of-memory issues

3.4 Hyperparameter

Cellpose

epochs	100
Optimizer	AdamW
Learning_rate	1e-5
weight_decay	0.1
batch_size	1

Detectron2

Batch_size	2
Learning_rate	5e-4
Batch_size_per_image	512
Score_Thresh_Test	0.5

3.5 Test-time Augmentation

We apply 7 test-time augmentation for cellpose model, and for Detectron2 model, we only do original image and vertical flip(1, 3)

1. Original image
2. Horizontal Flip
3. Vertical Flip
4. Horizontal Flip + Vertical Flip
5. Rotate 90 degree
6. Rotate 180 degree
7. Rotate 270 degree

3.6 Inference

Cellpose Inference Strategy

Probability and Flow Map Generation

- Generate probability maps indicating cell likelihood at each pixel
- Predict flow fields (horizontal and vertical gradients) pointing toward cell centers
- Apply 7-fold TTA with proper inverse transformations for flow vector alignment

Instance Recovery Process

- Use the predicted flow fields to construct a dynamical system
- Perform iterative gradient following from each pixel until convergence
- Group pixels converging to the same fixed point as single cell instances
- Apply flow threshold of 0.4 and minimum size filtering of 20 pixels

Detectron2 Inference Strategy

Multi-scale Patch-based Processing

- Process images at 6 different scales (440, 480, 520, 560, 580, 620 pixels)
- Split large images into overlapping patches (301×302 pixels) with stride (73×134) to prevent out-of-memory issues
- Retain only predictions with confidence scores ≥ 0.7
- Prioritize detections near patch centers to reduce edge artifacts

3.7 Ensemble

Ensemble Fusion Strategy

Combine predictions from Cellpose models and Detectron2 models

Overlap Resolution

- Use IoU threshold=0.15 to identify overlapping instances
- Apply weighted fusion based on prediction confidence scores
- Ensure no pixel is assigned to multiple instances in final output

4. Experimental Results

4.1 Cellpose Results

Figure 3 shows the batch visualization. Shows the four critical stages from cell probability prediction to final instance segmentation.

Figure 4 shows the flow field visualization result. Decomposes flow fields into x, y components and overall magnitude, proving the effectiveness of the technical approach.

Figure 5 shows the final result of Cellpose. Compares original images, ground truth annotations, and predictions to validate model performance.

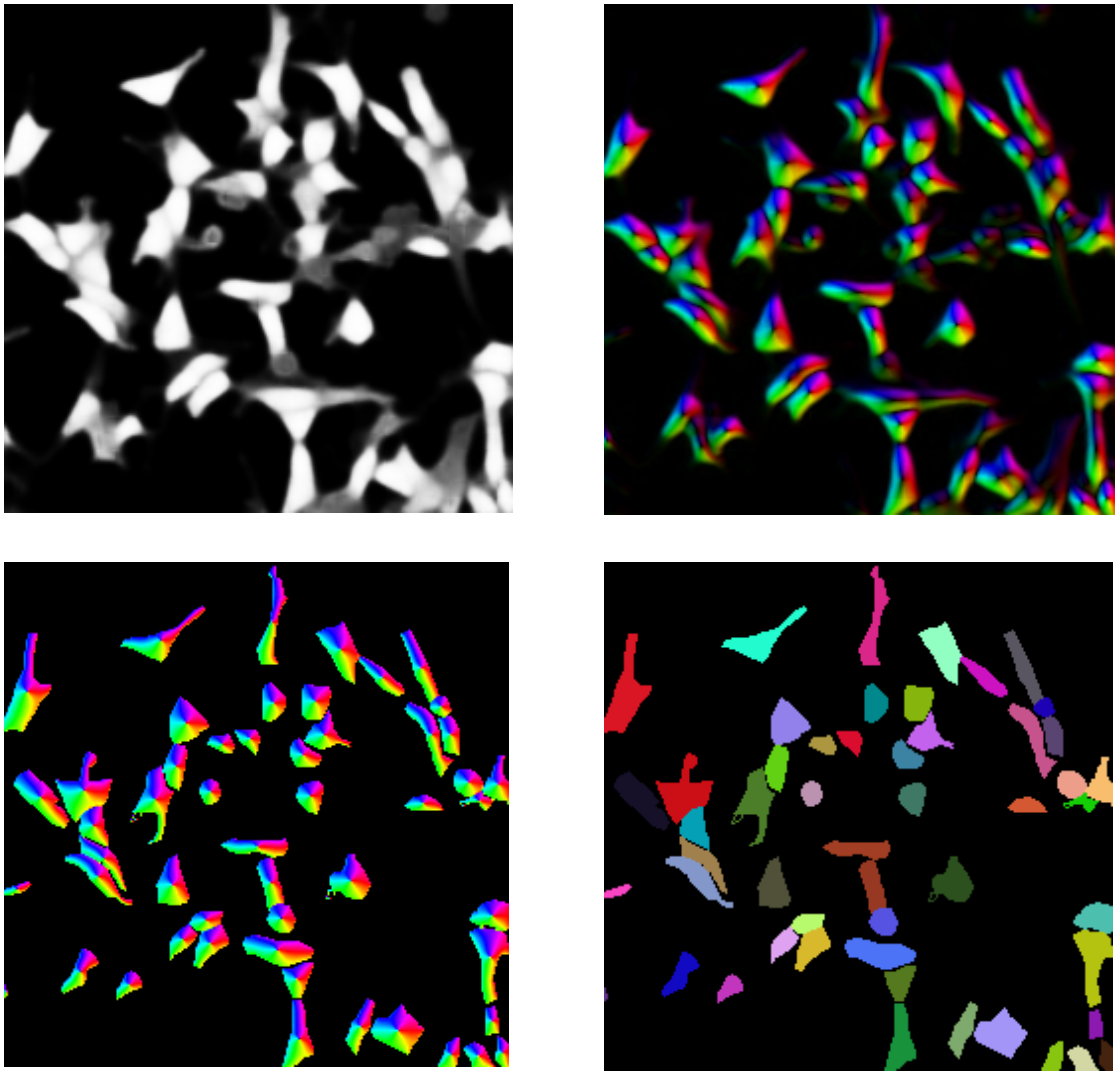


Fig.3 Batch Visualization Results
Predicted Cell Probability(Top Left), Predicted Flow(Top Right), Ground Truth Flow(Bottom Left), Ground Truth Mask(Bottom Right)

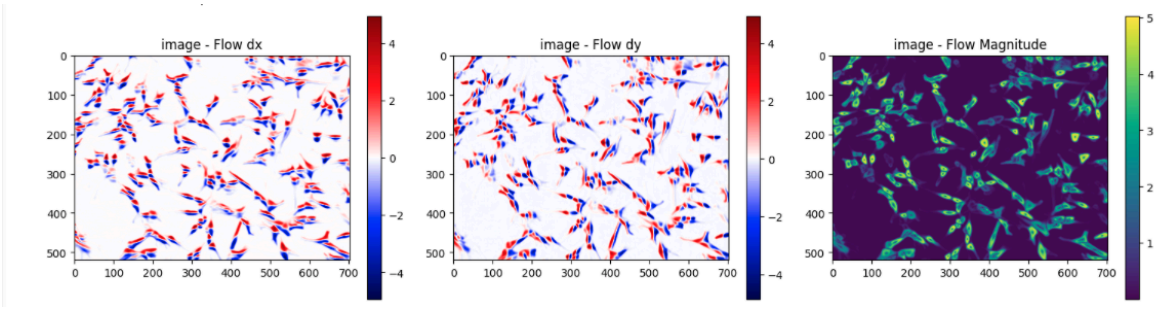


Fig.4 Flow Field Visualization Results

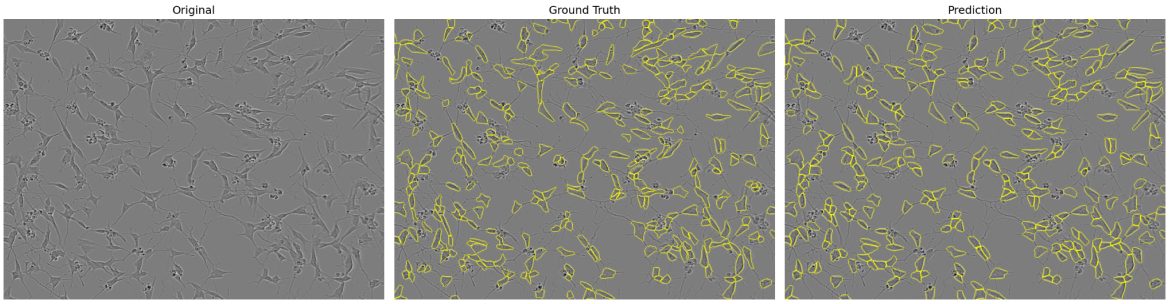


Fig.5 Prediction Result Comparison

4.2 Detectrons Results

Figure 6 shows the comparison between with and without Patch+TTA during testing. We found that these augmentation techniques help in detail.

Figure 7 shows the training convergence behavior of the Detectron2 model through loss curves over 10,000 iterations

non patch

patch + TTA

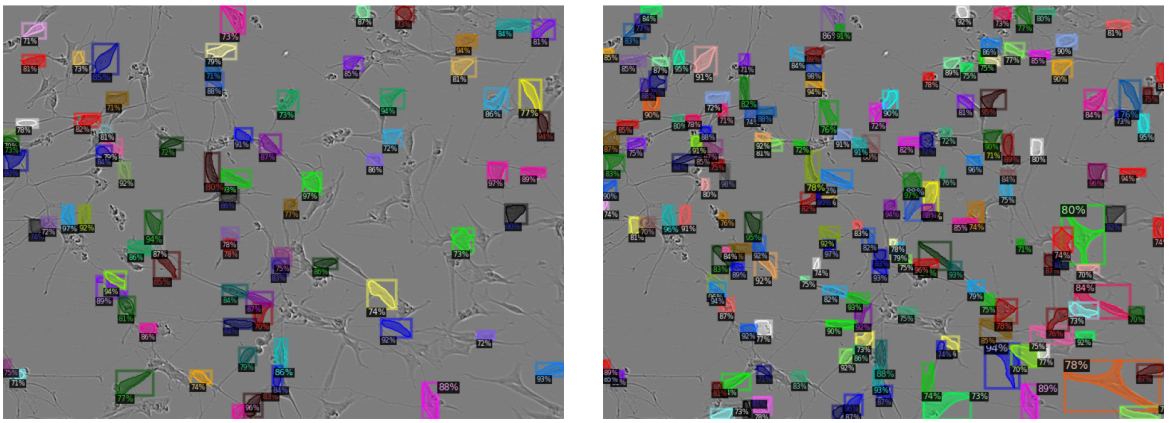


Fig.6 Patch+TTA Comparison

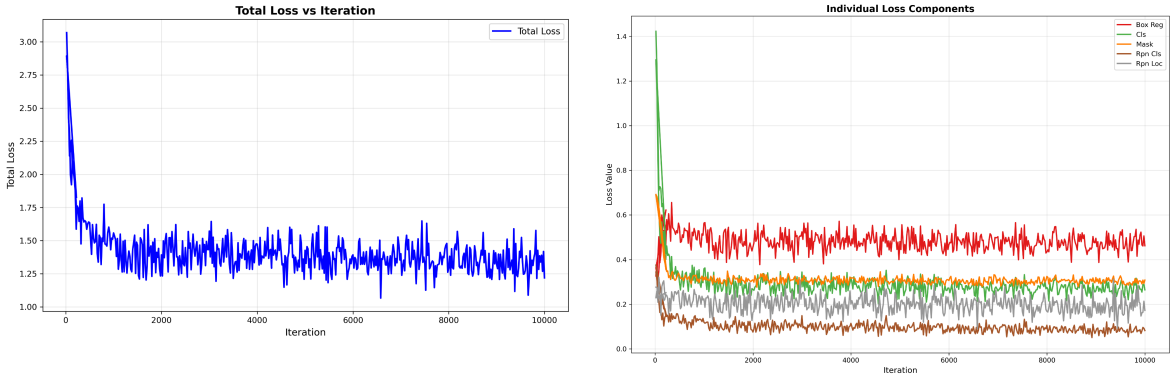


Fig.7 Loss/Iteration

4.3 Final Results

Figure 8 are the final results by ensemble model on testing data.

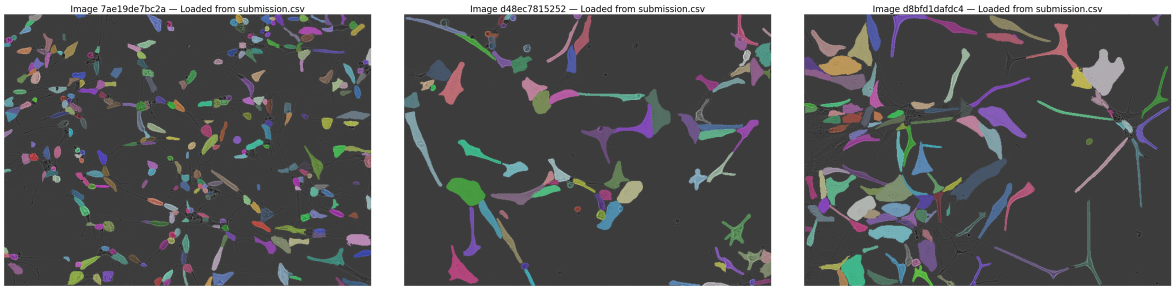


Fig.8 Final Result on Testing data

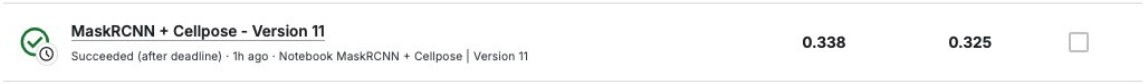


Fig.9 Kaggle competition results

5. Conclusion

This study proposes a neuronal cell instance segmentation method based on ensemble strategies. In our work, we experimented with two different models: Cellpose and Detectron2.

Technical Characteristics of Both Models:

Cellpose is a flow field-based model that computes gradient vector fields in x and y directions. Pixels belonging to the same cell converge to the same fixed point, achieving precise instance segmentation for irregular cell shapes. We adopted Detectron2's Mask R-CNN, which performs object detection via Region Proposal Networks (RPN) followed by pixel-level segmentation. This approach excels at detecting multiple separate objects in densely distributed scenarios.

Inference Strategies:

We applied different inference methods optimized through 5-fold cross-validation. For Cellpose, we used sliding window strategy and test-time augmentation. For Detectron2, we implemented multi-scale processing with confidence score filtering.

Ensemble Fusion Solutions:

We addressed three key challenges: First, output format differences were resolved through a unified post-processing pipeline. Second, we used IoU overlap as a consistency metric with dynamic weight allocation - enhancing weights for high overlap (IoU>0.15) and applying confidence-based weighting for significant differences. Third, we achieved precise

boundary fusion by combining Cellpose's accurate boundaries with Detectron2's stable detection through weighted mask fusion.

Experimental Results:

Our ensemble method achieved 0.338 mAP on the private dataset and 0.325 mAP on the public dataset, demonstrating good performance on cellular dataset. However, this result only reached the silver medal level in the competition. There is still room for improvement in the future work.

6. Reference

[1]. Stringer, C., Wang, T., Michaelos, M., & Pachitariu, M. (2021). Cellpose: a generalist algorithm for cellular segmentation. *Nature methods*, *18*(1), 100-106.

[2]. Pachitariu, M., Rariden, M., & Stringer, C. (2025). Cellpose-SAM: superhuman generalization for cellular segmentation. *bioRxiv*, 2025-04.

[3]. https://github.com/facebookresearch/detectron2

[4]. Ghiasi, G., Cui, Y., Srinivas, A., Qian, R., Lin, T. Y., Cubuk, E. D., ... & Zoph, B. (2021). Simple copy-paste is a strong data augmentation method for instance segmentation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 2918-2928).

[5]. https://cellpose.readthedocs.io/en/latest/models.html

[6]. He, K., Gkioxari, G., Dollár, P., & Girshick, R. (2017). Mask r-cnn. In *Proceedings of the IEEE international conference on computer vision* (pp. 2961-2969).

7. Team member contribution table

	313553044	313551139	313553046	313553037
Literature survey	15	25	30	30
Approach design	40	20	25	15
Approach implementation (experiment)	40	20	20	20
Report writing	10	35	20	35
Slide making and oral presentation	20	25	30	25