

Neuronal Cell Instance Segmentation

Keyu Nie, Stefan Zdrnja, Swathi Gangaraju
[keyunie, szdrnja, swathig1]@stanford.edu



Motivation

Accurate **instance segmentation of neuronal cells in microscopic images** could help understand how these cells interact with drugs. However, neuronal cells' distinct, irregular, and concave appearance makes this a complicated problem.

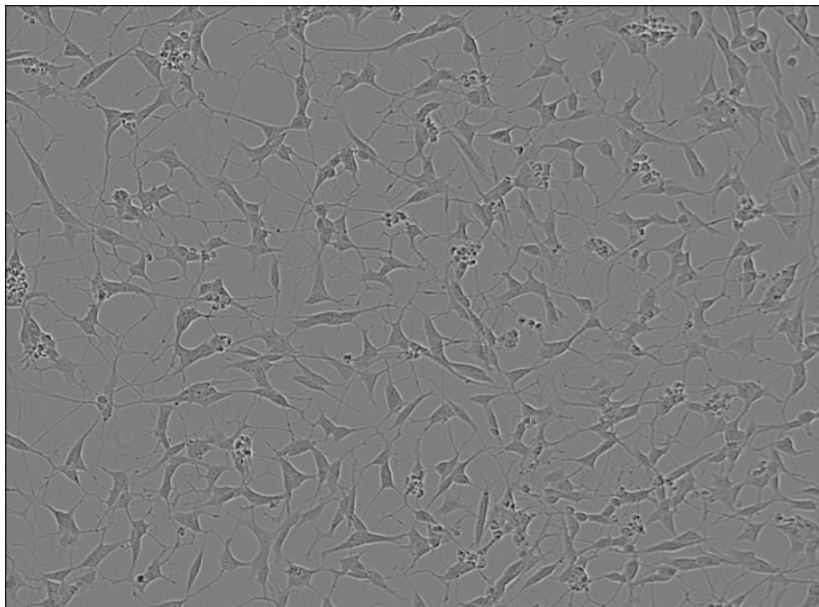


Figure 1. SHSY5Y cell type.

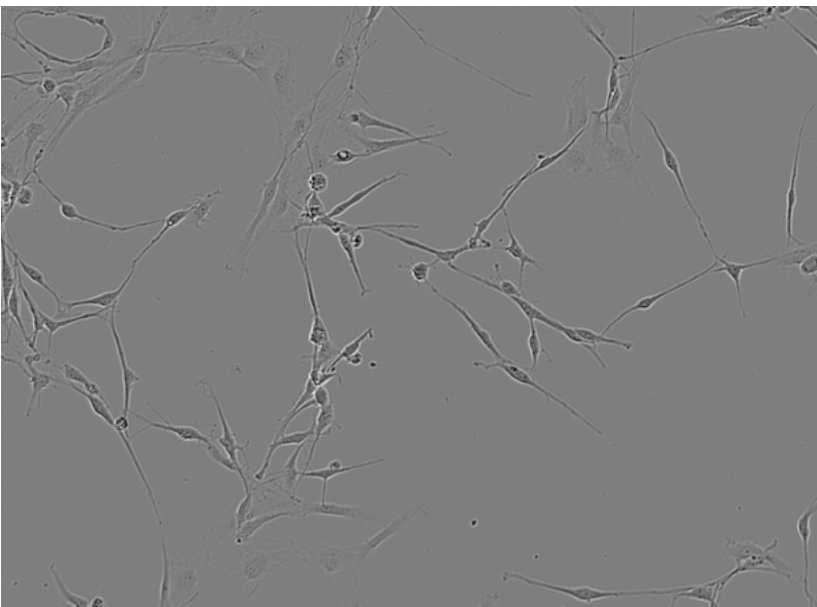


Figure 2. ASTRO cell type.

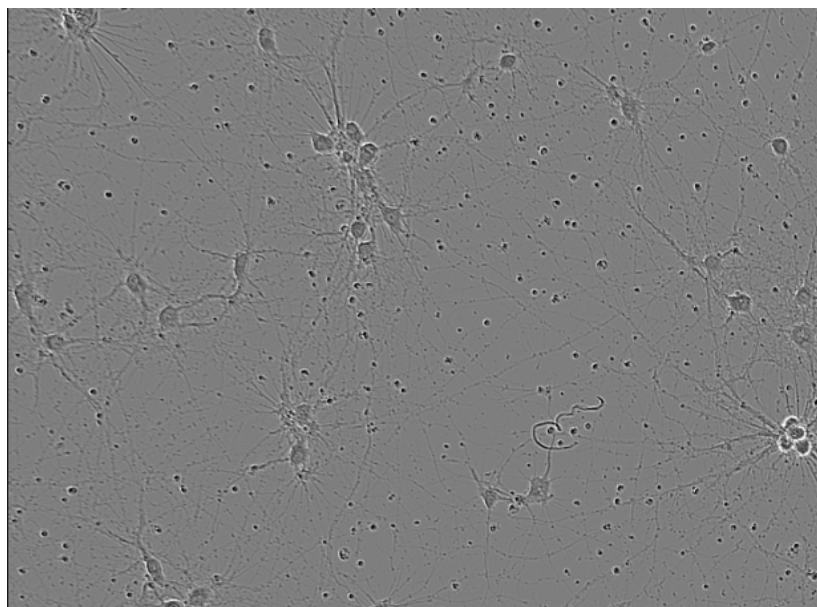


Figure 3. CORT cell type.

Dataset

Kaggle Sartorius Cell Instance Segmentation Challenge:

- 606 images for 3 cell types with 73585 annotations in COCO data format (140M).
- 52286 SHSY5Y (neuroblastoma), 10522 ASTRO (astrocytes), and 10777 CORT (neurons).
- Avg. Instance Size:** SHSY5Y (18, 18), ASTRO (51, 51), CORT (17, 17).
- Instance Density:** SHY5Y 50-450, ASTRO 1-200, CORT 15-50

Pretrain on LiveCell Dataset (Optional):

- 5, 239 images for 8 cell types with 1, 686, 352 annotations in COCO data format (1.3G).

Number of instances per image

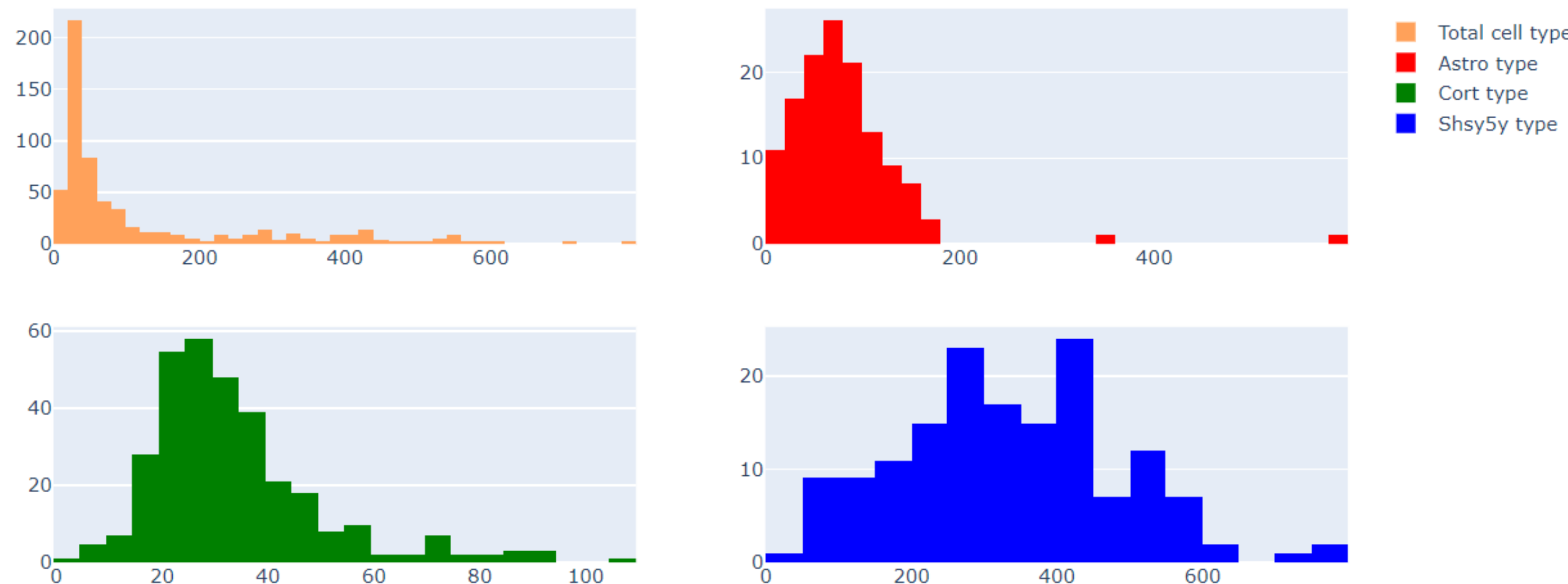


Figure 4. Dataset exploratory analysis.

Evaluation:

- Evaluation results via mean average precision (**mAP**) at different intersection over union (IoU) thresholds (IoU 0.5 : 0.05 : 0.95).

$$mAP = \frac{1}{\#thresholds} \sum_t \frac{TP(t)}{TP(t) + FP(t) + FN(t)}$$

Methods: Two stage architecture

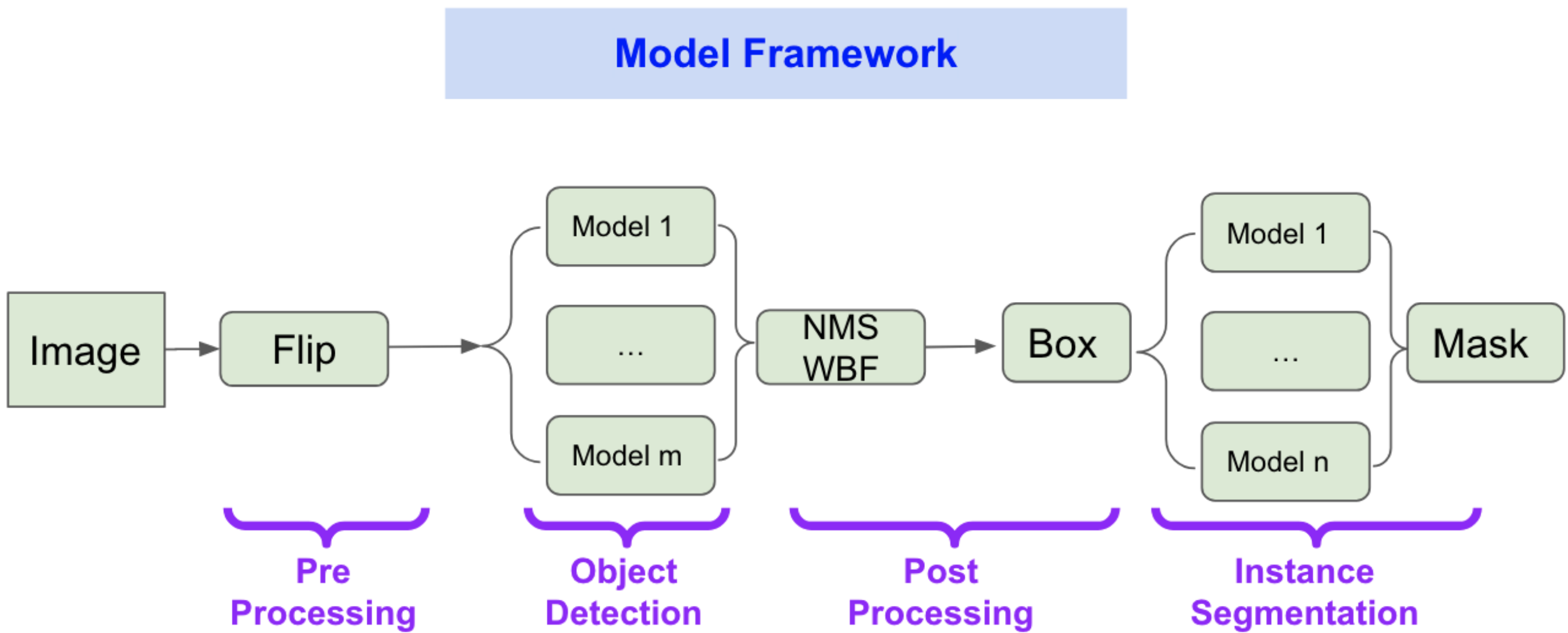


Figure 5. The framework of our models.

- Pre Processing:** RandomFlip, Resize, Normalization, Padding, Mosaic, MixUp, Photometric Distortion, Box Jitter.
- Object Detection:** Faster R-CNN, MASK R-CNN, RetinaNet, YoloX, SWIN-T(MASK R-CNN) (SWIN-Transformer as backbone in MASK R-CNN) and SWIN-T(RetinaNet) (SWIN-Transformer as backbone in RetinaNet)
- Post Processing:** NMS/WBF remove duplicate BBOXs, ROI-Alignment to feed one BOX per time to segmentor.
- Instance Segmentation:** MASK R-CNN, UperNet-SWIN-T (SWIN-Transformer as backbone encoder in UperNet structure)

Experiment Results

Model Comparison: YOLOX-UperNet-SWIN-T (Kaggle 1st solution, YOLOX as detector and UperNet-SWIN-T as segmentor) yields the best in our model search.

Model Name	Detector Backbone	Pretrained on Livecell Dataset	BBOX				
			mAP	AP by Category			COCO Bechmark
				shsy5y	astro	cort	
Mask R-CNN	ResNet50	N	0.266	0.135	0.262	0.401	0.382
Mask R-CNN	ResNet50	Y	0.278	0.133	0.292	0.408	0.382
Mask R-CNN-2	ResNet101	N	0.274	0.142	0.282	0.399	0.428
SWIN-T(Mask R-CNN)	SWIN-T	N	0.226	0.134	0.148	0.398	0.587
RetinaNet-UperNet-SWIN-T	ResNet50	N	0.216	0.076	0.202	0.372	N/A
YOLOX-UperNet-SWIN-T	YOLOX	Y	0.373 (0.396*)	0.317 (0.334*)	0.363 (0.393*)	0.439 (0.456*)	0.509

Figure 6. Model Detectors Performance Comparison.

Model Name	Segmentor Backbone	Pretrained on Livecell Dataset	Segmentation				
			mAP	AP by Category			COCO Benchmark
				shsy5y	astro	cort	
Mask R-CNN	ResNet50	N	0.247	0.149	0.181	0.413	0.347
Mask R-CNN	ResNet50	Y	0.258	0.150	0.207	0.419	0.347
Mask R-CNN-2	ResNet101	N	0.251	0.155	0.187	0.411	0.384
SWIN-T(Mask R-CNN)	SWIN-T	N	0.205	0.141	0.079	0.396	0.511
RetinaNet-UperNet-SWIN-T	UperNet-SWIN-T	Y	0.204	0.076	0.146	0.387	N/A
YOLOX-UperNet-SWIN-T	UperNet-SWIN-T	Y	0.337 (0.362*)	0.310 (0.328*)	0.261 (0.302*)	0.440 (0.456*)	0.502

Figure 7. Model Segmentors Performance Comparison.

Figure 8. Model Comparison. The numbers with * marker are reported in Kaggle first place authors threads.

Results and Discussion

Ablation Study of the Best Model YOLOX-UperNet-SWIN-T:

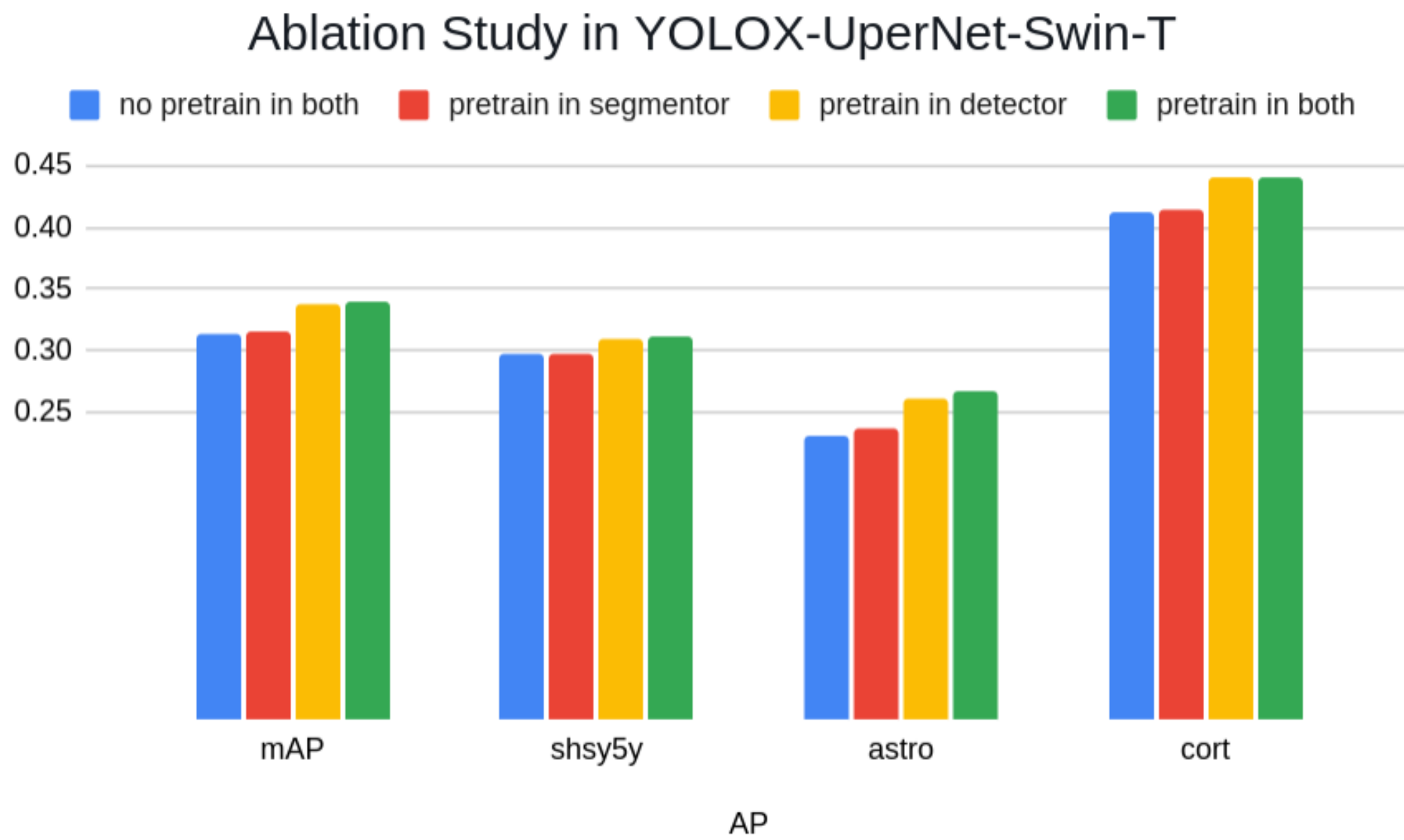


Figure 9. Performance analysis on YOLOX-UperNet-Swin-T model with/o pretrained on LiveCell data for both object detector and instance segmentation.

Lesson learned:

- Resize to bigger shape would improve model performance.
- Pretrain on LiveCell dataset increase model performance, especially for **object detector**.
- YOLOX is the key for the success of **YOLOX-UperNet-SWIN-T**
- WBF shows potential for model ensemble, but we need better models.
- Memory issue when using larger batch size or bigger dataset.

Visualization for prediction from YOLOX-UperNet-SWIN-T:

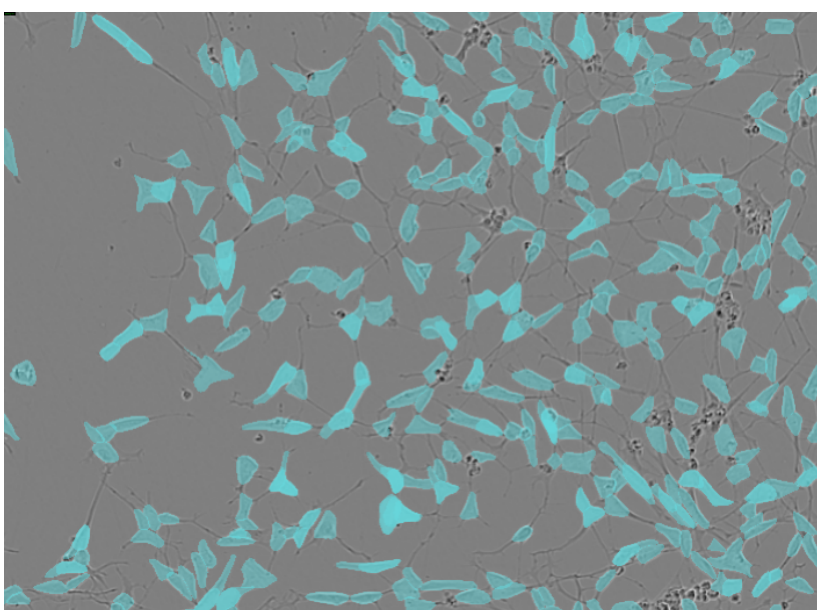
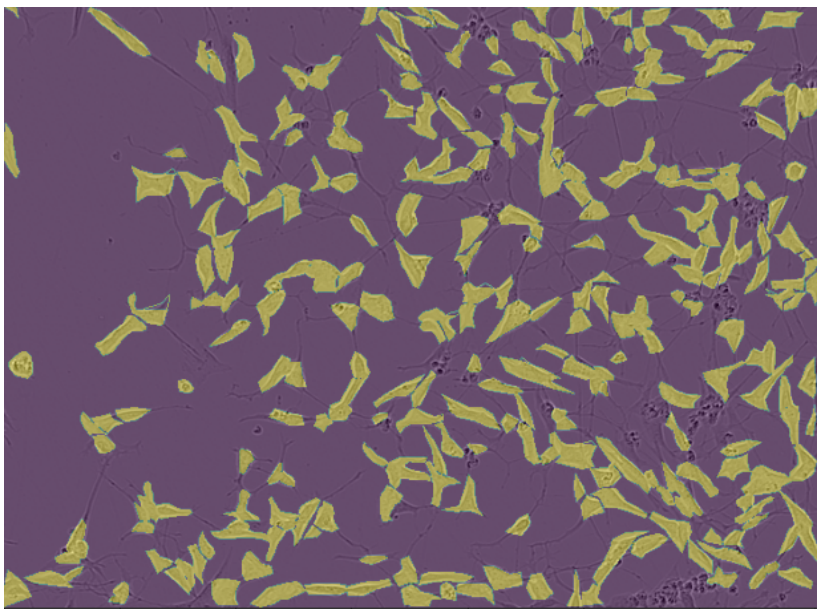


Figure 10. SHSY5Y ground truth (above) vs prediction (below).

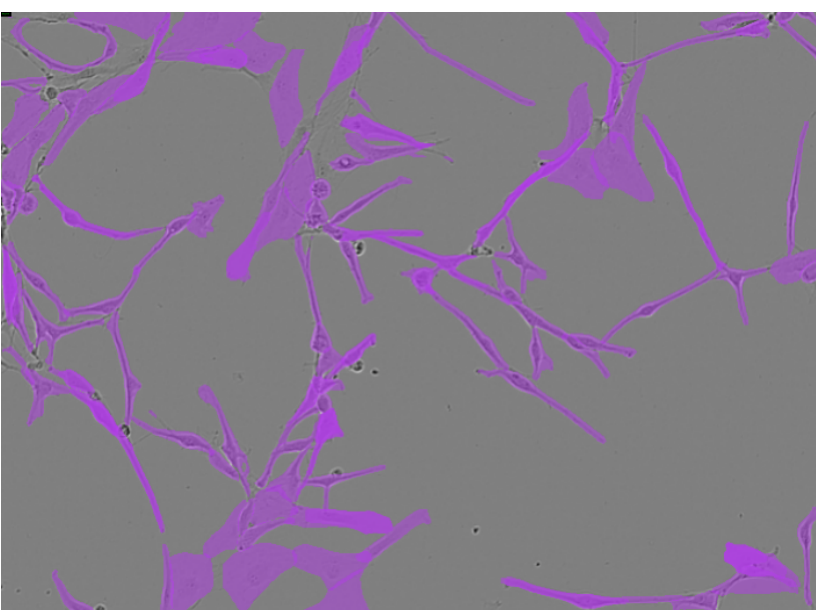
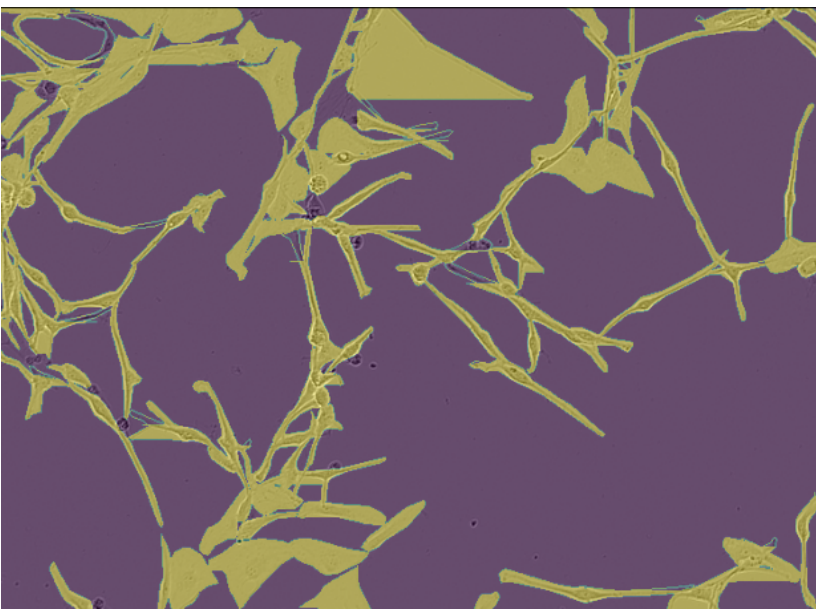


Figure 11. ASTRO ground truth (above) vs prediction (below).

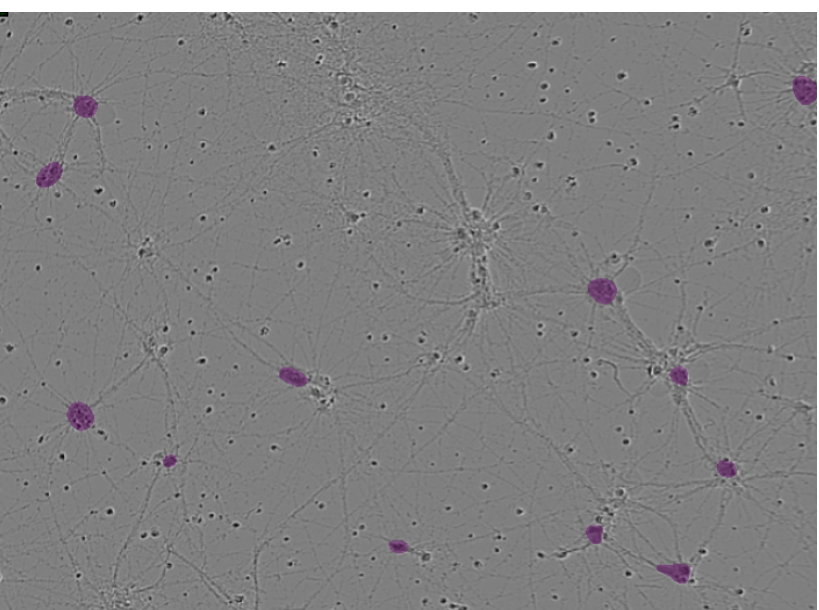
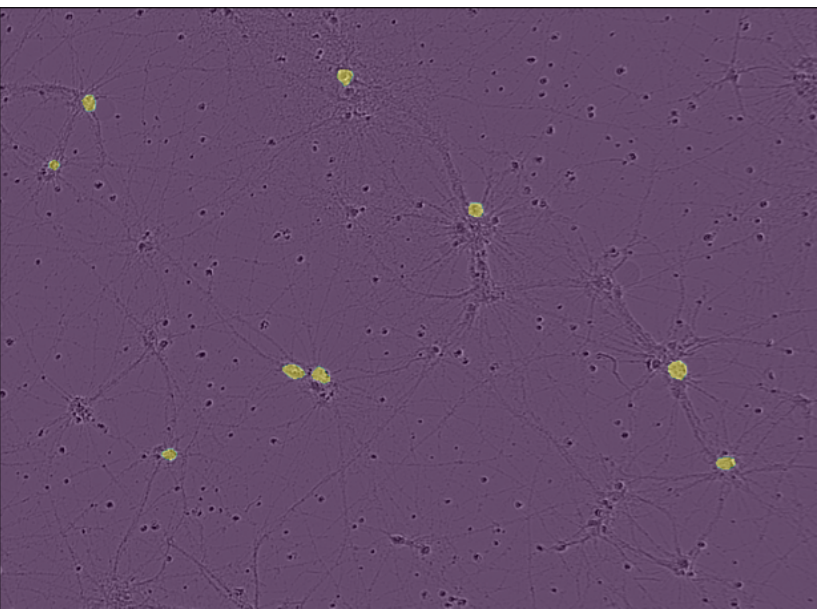


Figure 12. CORT ground truth (above) vs prediction (below).

Conclusion and Future Work

- YOLOX-UperNet-SWIN-T yielded best results for this problem.
- We achieved a 0.373 mAP score on detection and mAP 0.337 score on segmentation.
- Next steps:** Increase the sample size given better memory and computing power. Explore other feature extraction techniques to learn based on variations in instances.