

精準蘆筍影像監測系統之植株識別

**Plant Identification of a Precision Asparagus Image
Monitoring System**

Reporter: Chiao-Chi Hsu

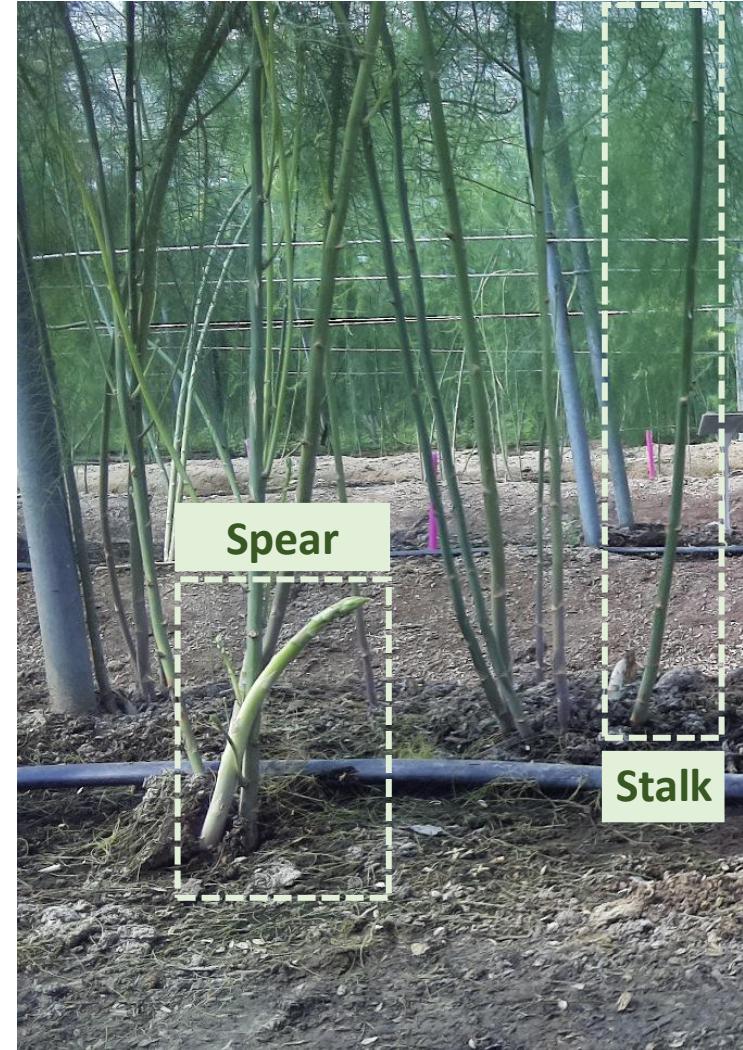
Advisor: Dr. Shih-Fang Chen

Introduction



Mother stalk method

- Keep a certain amount of stalks for photosynthesis
- May harvest the spears twice a day.
- **Weakness: Extra labor is required.**



Introduction

Asparagus Greenhouse Management System

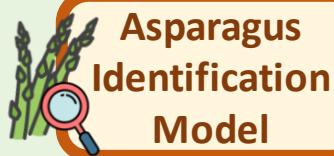
Previous

Field Robot

Self-guided System



Current



Density

Location

Future

Spraying Pesticides
and Fertilizers Vehicle

Objectives

- Build an Asparagus module using Masked-attention Mask Transformer (Mask2Former) on different dataset and compare it to the previous Mask R-CNN and Mask DINO model.

Materials and Methods

Dataset

- Asparagus growth images were collected from August 2019 to November 2023.
- The main shooting equipment was Raspberry Pi Camera v2.
- A total of **2246 images** were collected, labeled as **stalk** and **spear**.

# stalk in 1 meter	Density
< 15	Sparse
15~25	Medium
> 25	High

Total dataset	Original	Half	High Density
# Images	2212	1186	2212 + 34

Materials and Methods

Plant Density

Density definition

# stalk in 1 meter	Density
< 15	Sparse
15~25	Medium
> 25	High



Density : Sparse



Density : Medium

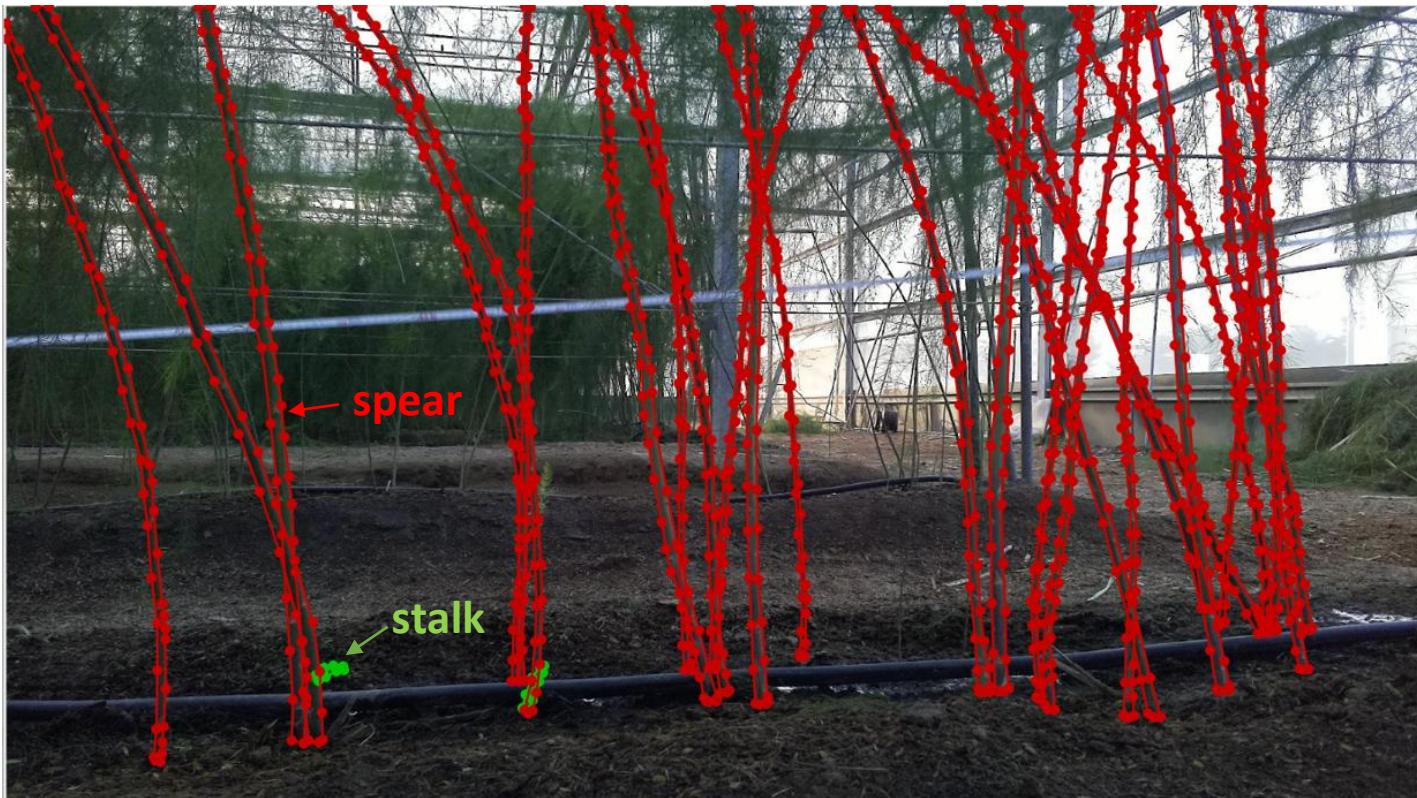


Density : High

Materials and Methods

Annotation of High Density Images

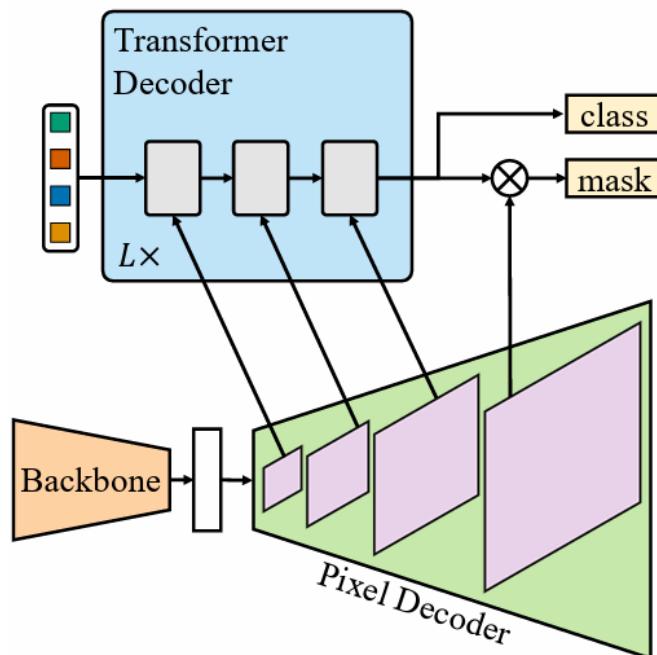
- High complexity of crossings



Materials and Methods

Module: Mask2Former overview

- Mask2Former is a **transformer-based instance segmentation model**.
- This model consists of a feature extraction backbone, a pixel-decoder, a transformer-decoder.



Mask2Former overview

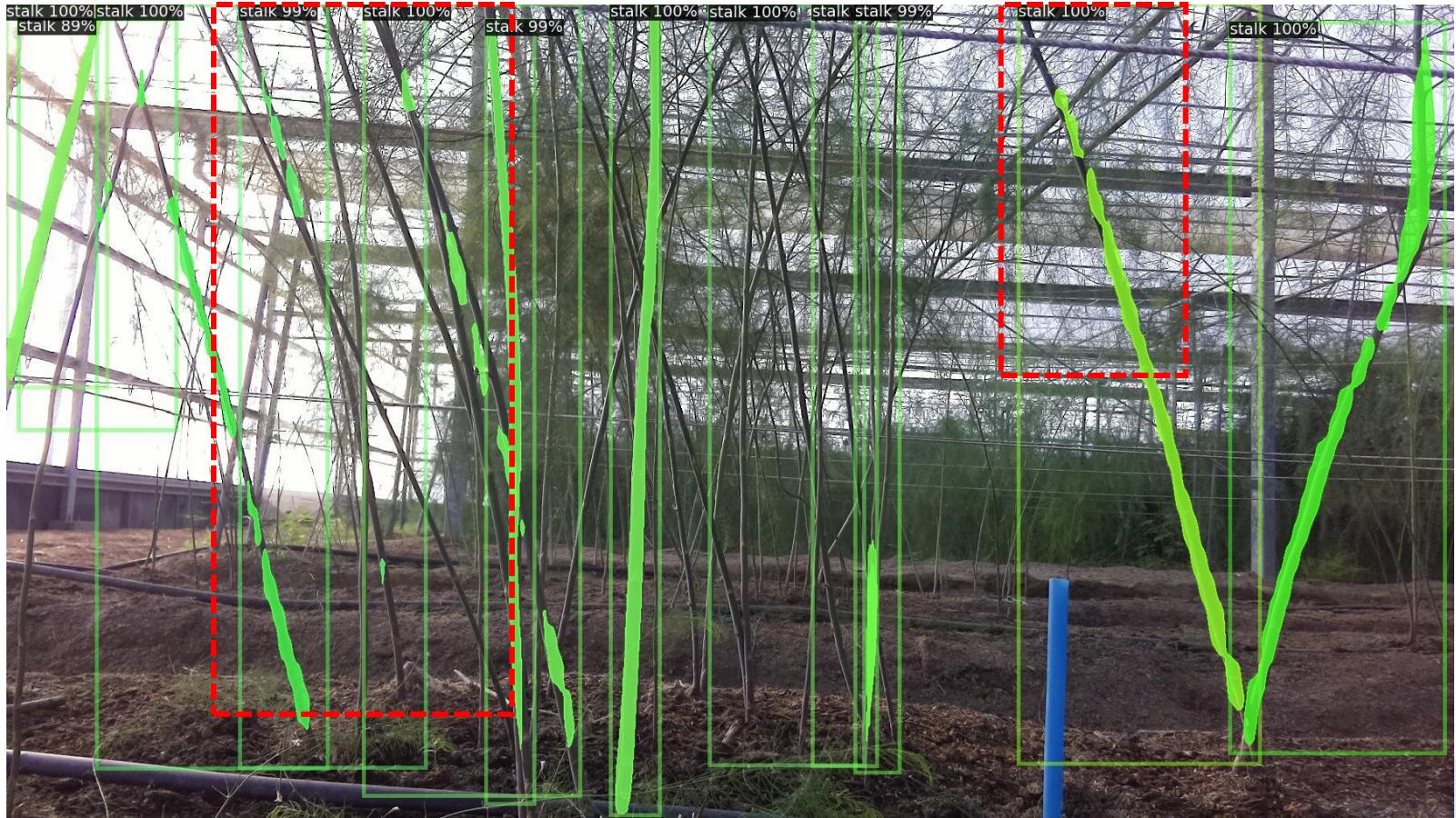
Result and Discussion

Model Performance

model	Dataset (images)	AP	AP50	AP50 of stalk	AP50 of spear
Mask R-CNN	Original	31.90	65.55	55.97	75.13
Mask DINO	Original	54.52	81.01	87.58	74.44
Mask2former	Original	48.36	83.74	83.04	80.63
Mask2former	Half	47.67	83.69	83.58	79.36
Mask2former	High Density	47.39	86.74	86.74	79.60

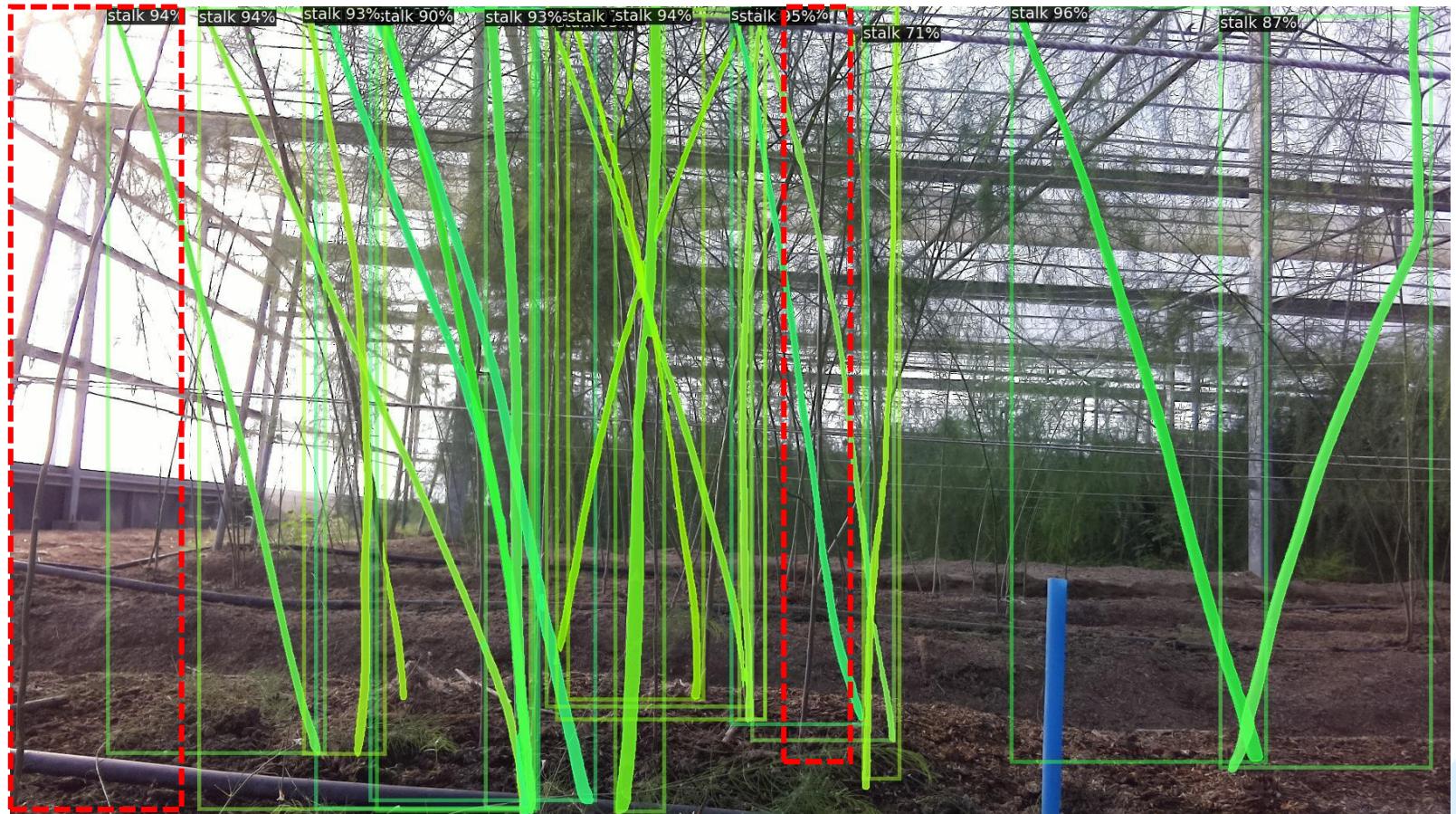
Result and Discussion

Prediction of Mask R-CNN



Result and Discussion

Prediction of Mask DINO



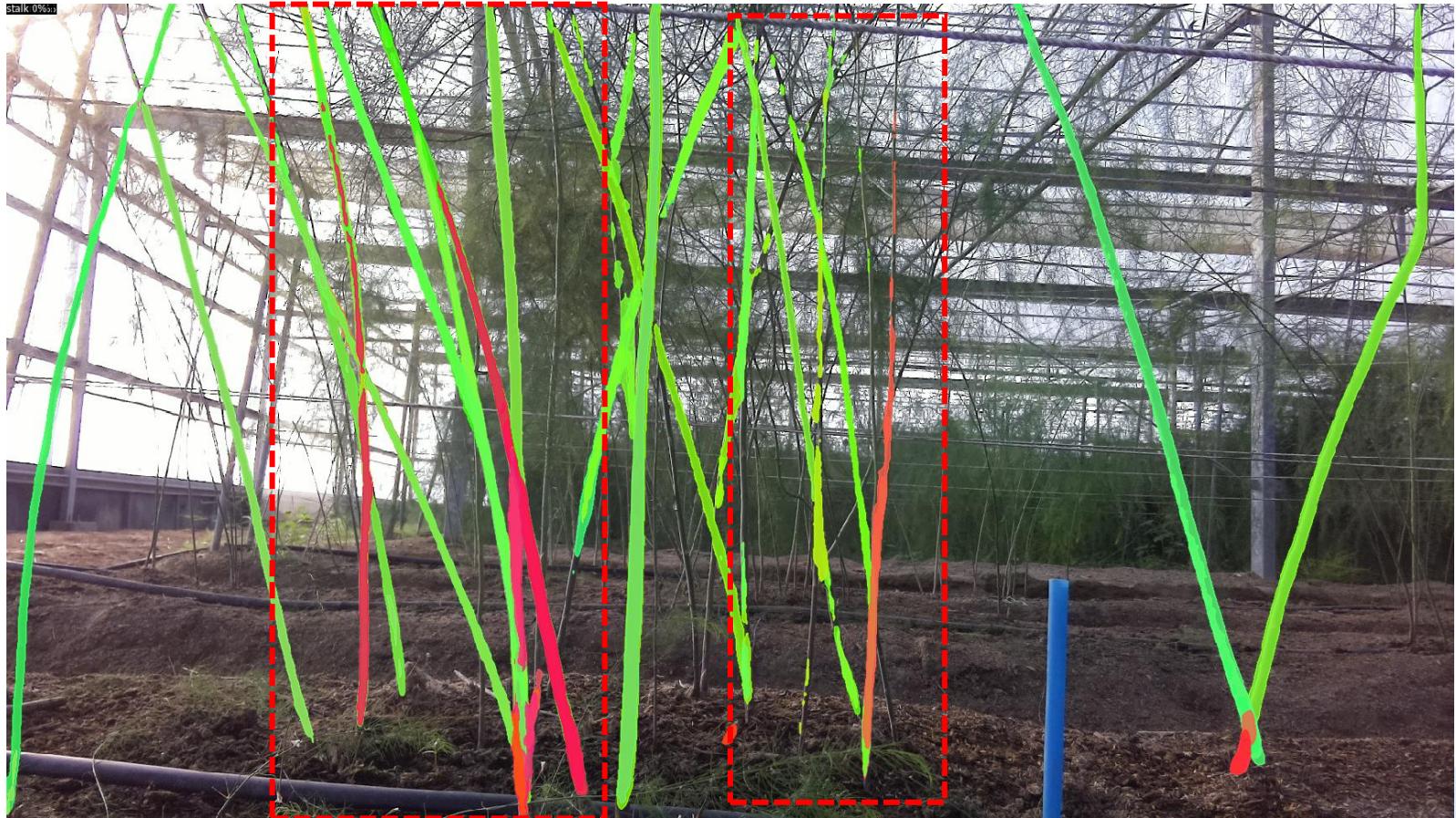
Result and Discussion

Prediction of Mask2former (Original dataset)



Result and Discussion

Prediction of Mask2former (Half dataset)



Result and Discussion

Prediction of Mask2former (High Density dataset)



Conclusion

- This research contrasts different models and dataset of identifying asparagus.
- Within the original dataset, the performance of the Mask2former model is close to the Mask DINO model and better than the Mask R-CNN model.
- On Mask2former, with the Half and High Density dataset, its performance is similar to Mask DINO, however it improves the mask prediction on thinner stalks.

Future Work

- Based on Half dataset, adding high density images in training dataset.

Reference

- Cheng, B., Misra, I., Schwing, A. G., Kirillov, A., & Girdhar, R. (2022). Masked-attention mask transformer for universal image segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 1290-1299).
- Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., ... & Guo, B. (2021). In *Proceedings of the IEEE/CVF international conference on computer vision* (pp. 10012-10022).
- 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).
- Li, F., Zhang, H., Xu, H., Liu, S., Zhang, L., Ni, L. M., & Shum, H. Y. (2023). Mask DINO: Towards a unified transformer-based framework for object detection and segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 3041-3050).



Thanks for Your Attention