Dual-Model Fusion for Personal Protective Equipment Detection

Project 2

SoICT, HUST

Introduction

- PPE ensures workplace safety in construction, manufacturing
- Traditional PPE detection lacks accuracy in complex conditions

→ Build high-accuracy PPE detection using dual-model fusion

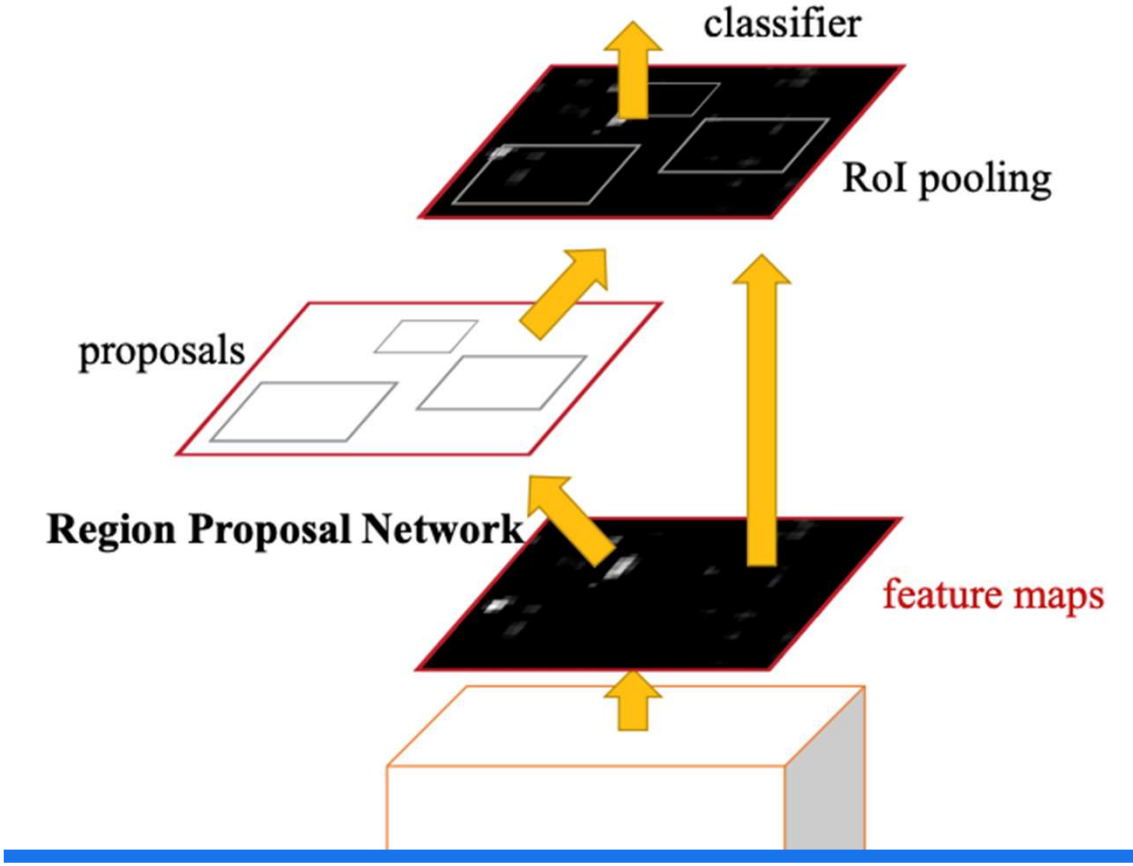
Background

Faster-RCNN [1]

- Faster R-CNN is a two-stage object detector that first generates region proposals and then classifies them

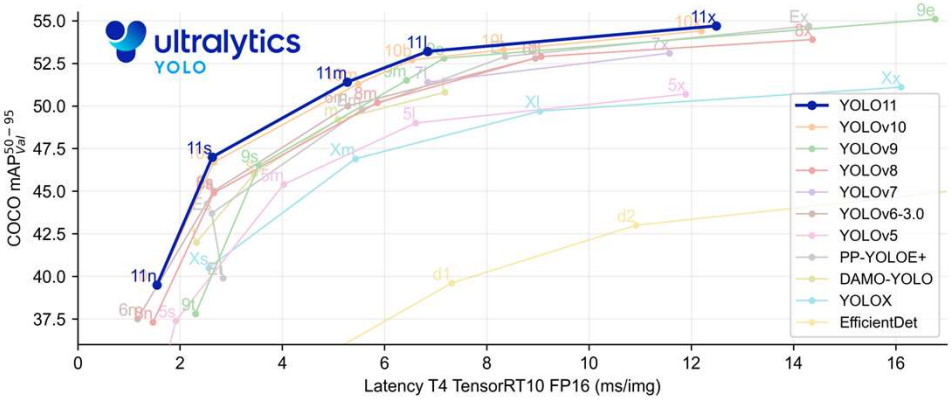
→the pipeline more computationally intensive and harder to optimize end-to-end.

→Slow inference speed



[1] **Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks** Ren, S., He, K., Girshick, R. and Sun, J., 2015.

Background

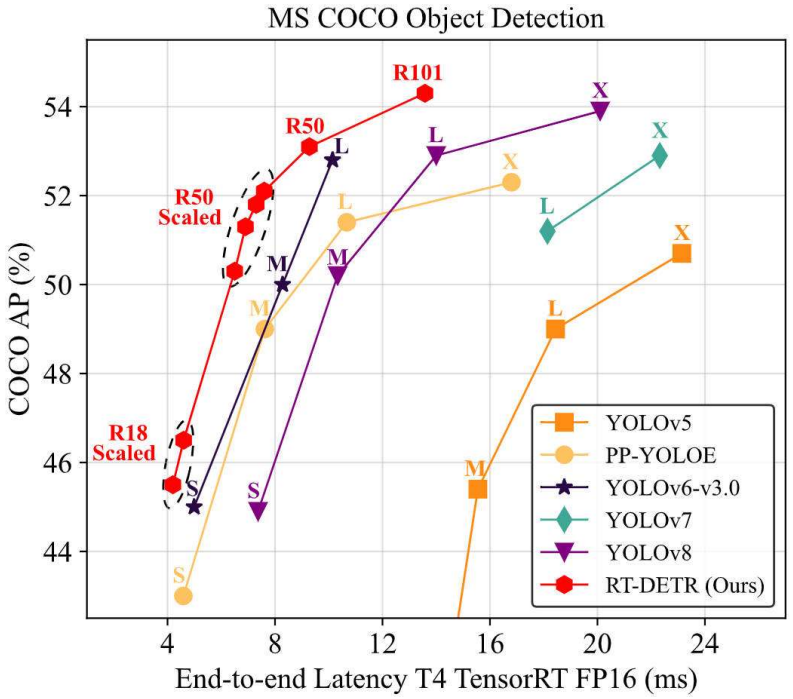


YOLO (You Look Only Once) [2,3]

- YOLO is a real-time object detection model that frames detection as a single regression problem
- offers a compelling balance between speed and accuracy
- **Limitations: Lower accuracy on small objects and dense scenes**

[2] Redmon, J. and Farhadi, A., 2018. YOLOv3: An incremental improvement. *arXiv preprint arXiv:1804.02767*.
[3] https://www.ultralytics.com/

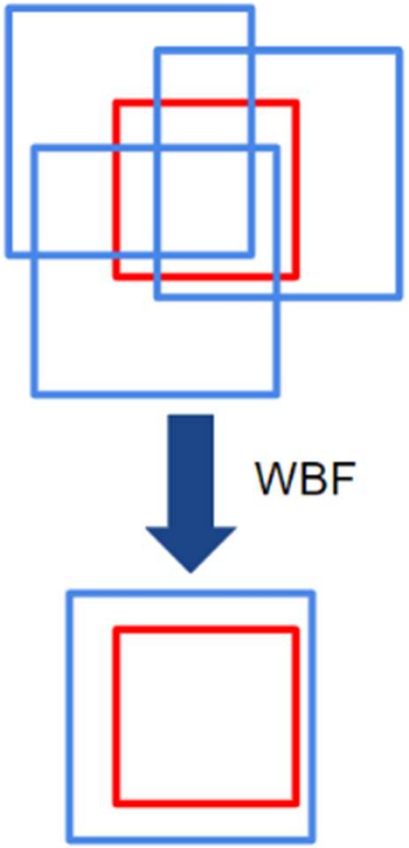
Background



RTDETR [4]

- RT-DETR is a transformer-based detection architecture designed for **real-time performance**, combining the benefits of end-to-end DETR models with optimizations that reduce latency
- Compared to YOLO : Better generalization, better for small objects.
- Currently support three versions: v1, v2, v3

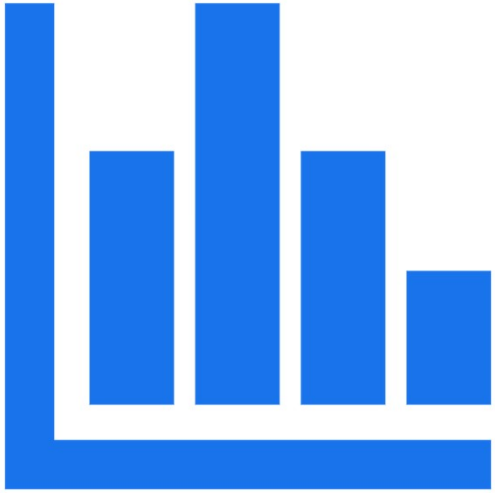
[4] Lv, W., Zhao, Y., Chang, Q., Huang, K., Wang, G. and Liu, Y., 2024. *RT-DETRv2: Improved Baseline with Bag-of-Freebies for Real-Time Detection Transformer*. arXiv preprint arXiv:2401.12140.



Dual-Model Fusion Overview

- Each model excels uniquely, YOLO: fast but not good for small objects ; RTDETR : higher accuracy for small objects.
- In this project, Weighted Boxes Fusion (WBF) [5] is used to combine bounding boxes from two models

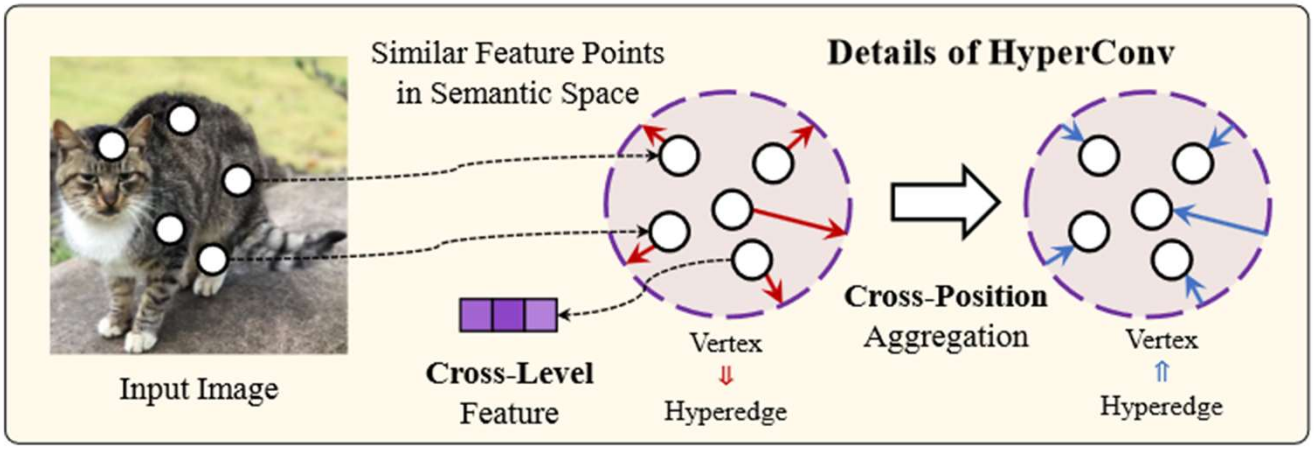
[5] Solovyev, R., Wang, W. and Gabruseva, T. (2021) 'Weighted boxes fusion: Ensembling boxes from different object detection models', Image and Vision Computing, 107, p. 104117. doi: 10.1016/j.imavis.2021.104117.



METHODOLOGY

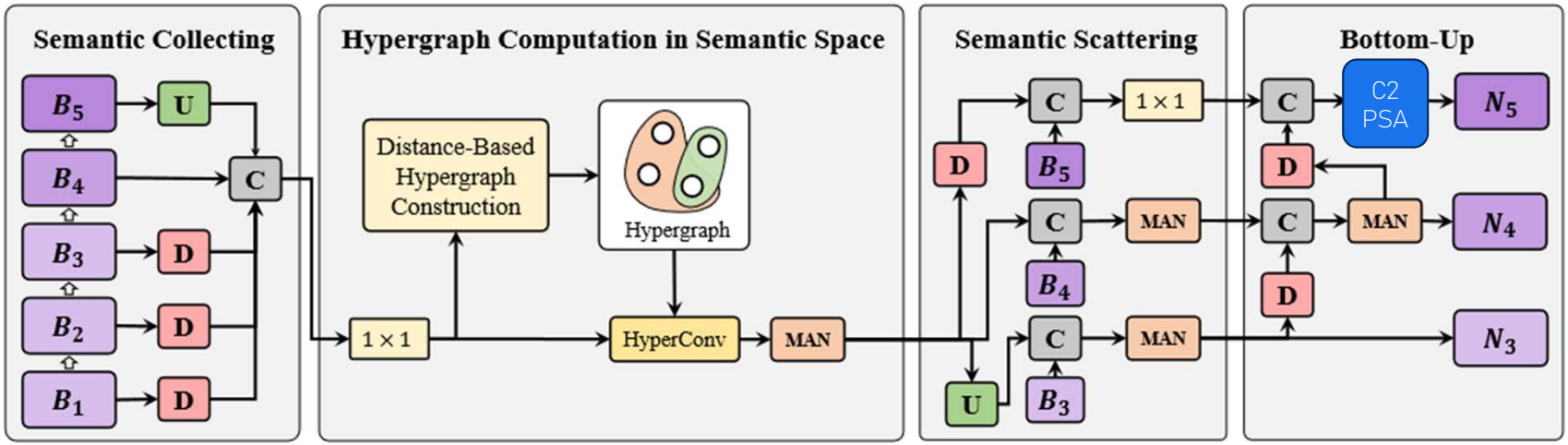
Hyper-YOLO

- Hyper-YOLO is a new object detection method that integrates hypergraph computations to capture the complex high-order correlations among visual features.

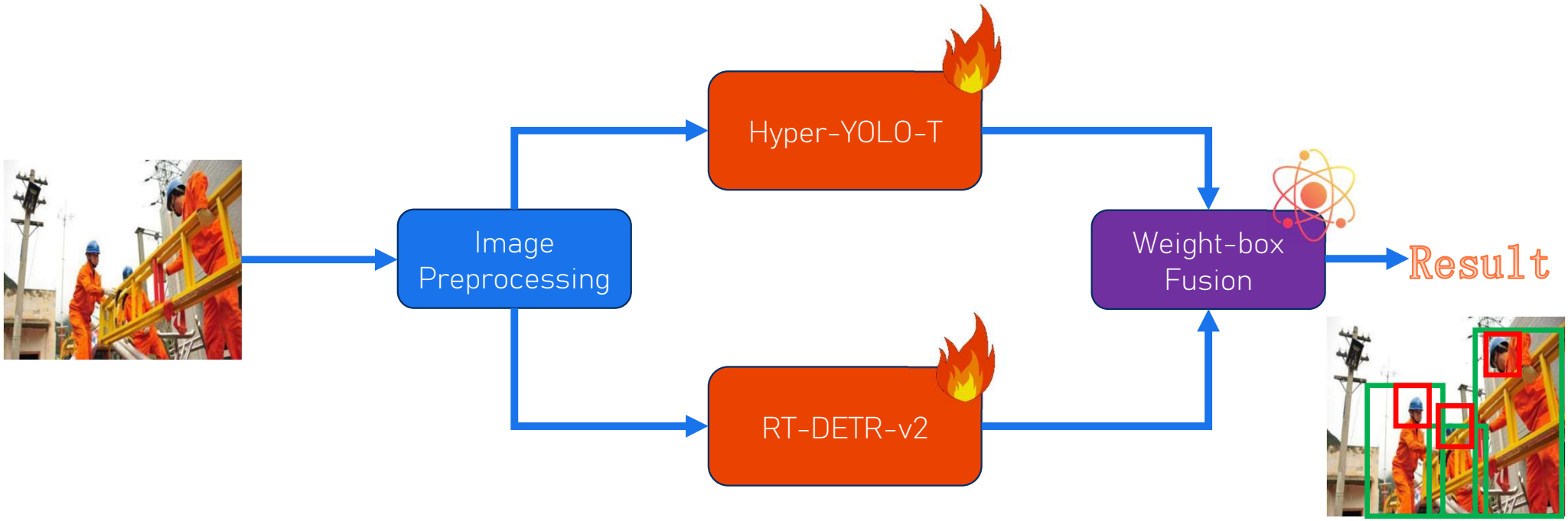


Proposed Model : Hyper-YOLO-T

To help the Hyper-YOLO backbone extract better features, we proposed the Hyper-YOLO-T to replace the C2f module with C2PSA (introduced in YOLOv11)



Pipeline


Hyper-YOLO-T
Image Preprocessing
Weight-box Fusion
Result
RT-DETR-v2

Dataset

- Dataset Source: PPE data obtained from Roboflow [6].
- Classes: 9 classes including boots, gloves, hardhat, no_boots, no_gloves, no_hardhat, no_vest, person, vest.

boots

hardhat

no_gloves

no_vest

TRAIN SET 70%	VALID SET 20%	TEST SET 10%
2898 Images	821 Images	416 Images

[6] https://universe.roboflow.com/ppe-yngjj/ppe-vum8g/dataset/10

Evaluation Metrics

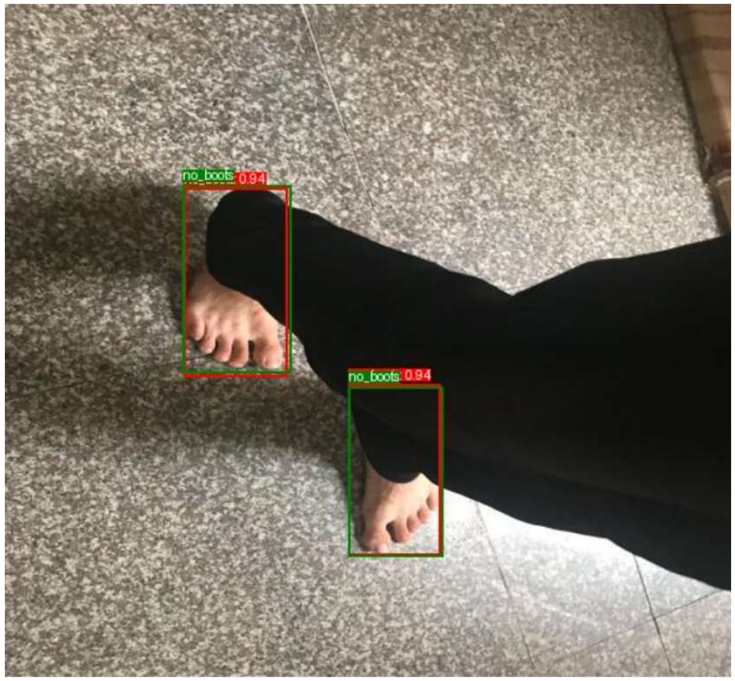
- **Precision**: Measures how accurate the model is when it makes a detection. It is the ratio of true positives (TP) to all predicted positives (TP + FP).
- **Recall**: Measures how well the model finds all relevant objects. It is the ratio of true positives (TP) to all actual objects (TP + FN).
- **mAP@0.5 (mAP50)**: Mean Average Precision at IoU threshold 0.5. A common metric for evaluating detection accuracy when allowing a 50% overlap between predicted and ground truth boxes.
- **mAP@0.5:0.95**: Averaged mAP over IoU thresholds from 0.5 to 0.95 with step 0.05. This COCO-standard metric offers a more comprehensive and strict evaluation of model performance.

Experimental Results

	Precision	Recall	mAP50	mAP50-95
YOLOv11(large version)	0.745	0.731	0.759	0.436
RT-DETRv2	0.748	0.735	0.750	0.433
YOLOv10(large version)	0.737	0.633	0.67	0.369
HyperYOLO	*0.765*	0.648	0.699	0.395
HyperYOLO-T (proposed)	0.762	**0.749**	*0.772*	*0.440*
Proposed Pipeline	**0.768**	*0.745*	**0.773**	**0.448**

Our pipeline reaches 0.773 mAP50 , the highest among current object detection models

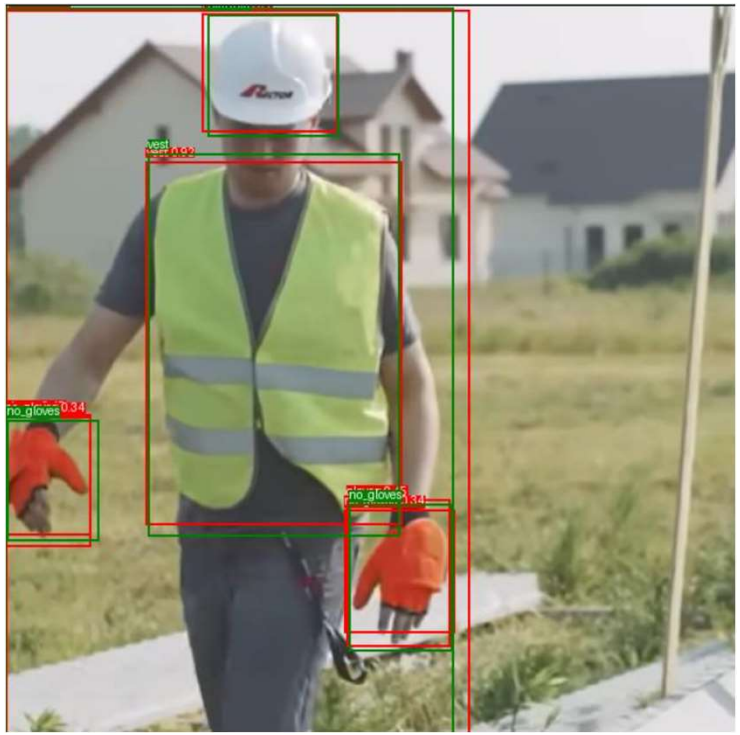
Examples



Ground-truth

Predicted

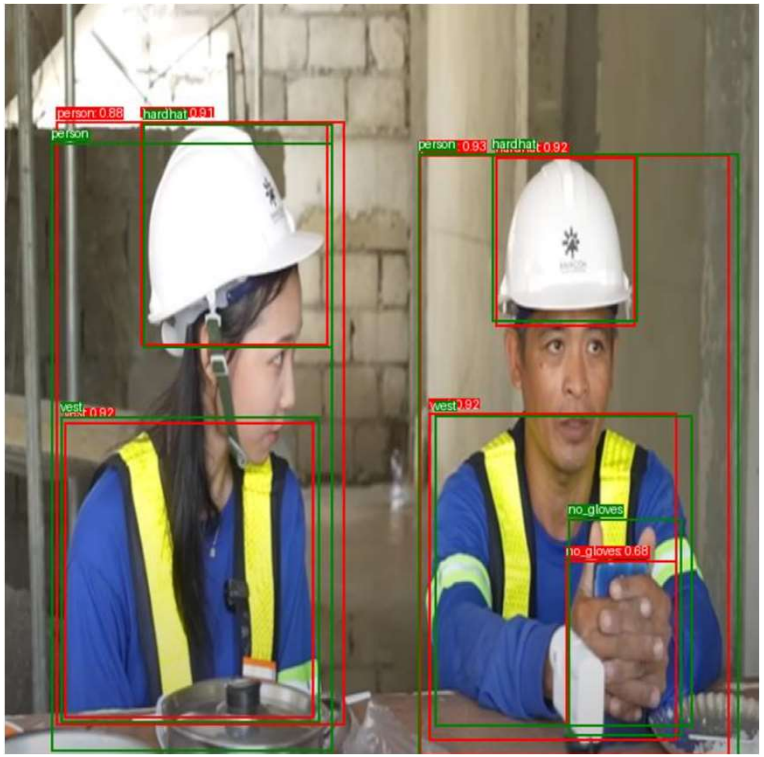
Examples



Ground-truth

Predicted

Examples



Ground-truth

Predicted


THANK YOU
FOR
LISTENING