YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors

Chien-Yao Wang, Alexey Bochkovskiy, and Hong-Yuan Mark Liao

Institute of Information Science, Academia Sinica, Taiwan

Poster: TUE-PM-319

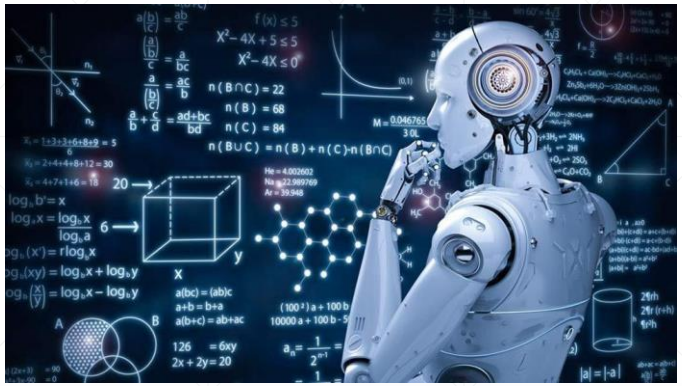
Presenter : Chien-Yao Wang

- **Motivation & Introduction**
- Proposed Method & Results
- Applications & Conclusion

Motivation

- New techniques solve a lot of existing problems. However they also bring new research topic to human.

New technologies



New issues

- We find several new research topics from novel object detection technologies, and propose new solutions to address them.

Image credit:
https://misti.mit.edu/misti-impact/impact-artificial-intelligence
https://www.digiconasia.net/tips-strategies/the-four-key-barriers-to-solving-great-problems-with-machine-learning

Introduction (1/2)

- In this work, we
  - 1. design an extended version of efficient layer aggregation network.
  - 2. analyze the model scaling factors for concatenation-based networks, and design a simple yet effective model scaling strategy.

Introduction (2/2)

- We also use trainable bag-of-freebies to solve
  - 1. make re-parameterization modules can work with modern networks.
  - 2. make lead head and auxiliary head can learn consistency information from dynamic label assignment methods.

Bag-of-freebies
1. Improve accuracy
2. May increase training cost
3. No additional inference cost

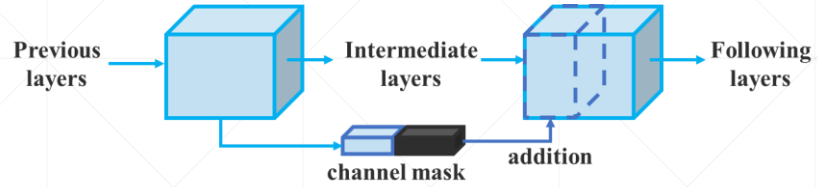
- Motivation & Introduction
- **Proposed Method & Results**
  - **Extended efficient layer aggregation networks**
  - **Model scaling for concatenation-based models**
  - **Deep supervision meets dynamic label assignment**
  - **Re-parameterization model meets modern networks**
  - **Results**
- Applications & Conclusion

Extended efficient layer aggregation networks

Optimize gradient path instead optimize data path

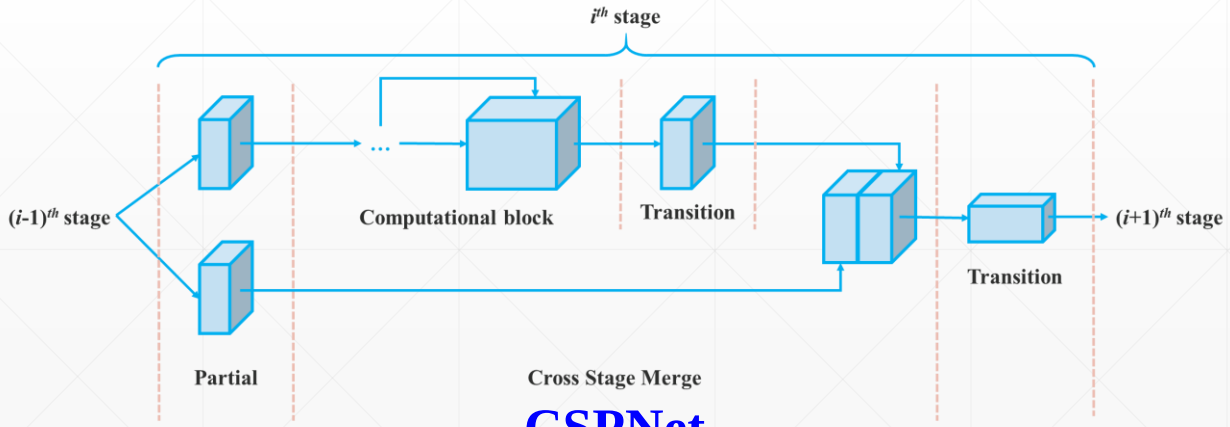
- **We design network architecture based on gradient path analysis.**

Layer-level design: partial residual network (PRN)



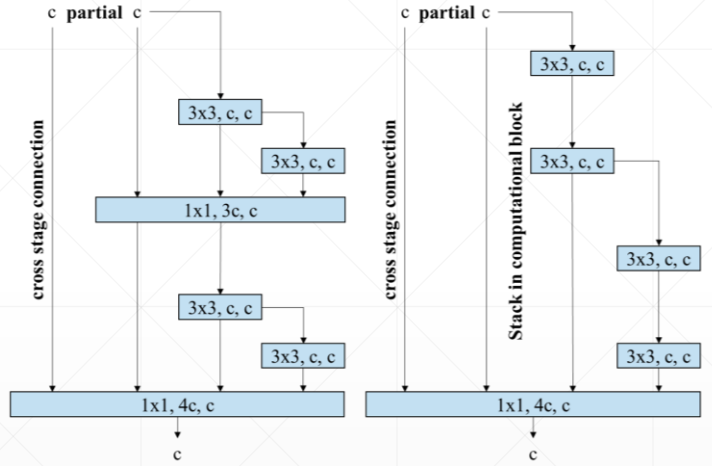
PRN

Stage-level design: cross stage partial network (CSPNet)



CSPNet

Network-level design: efficient layer aggregation network (ELAN)

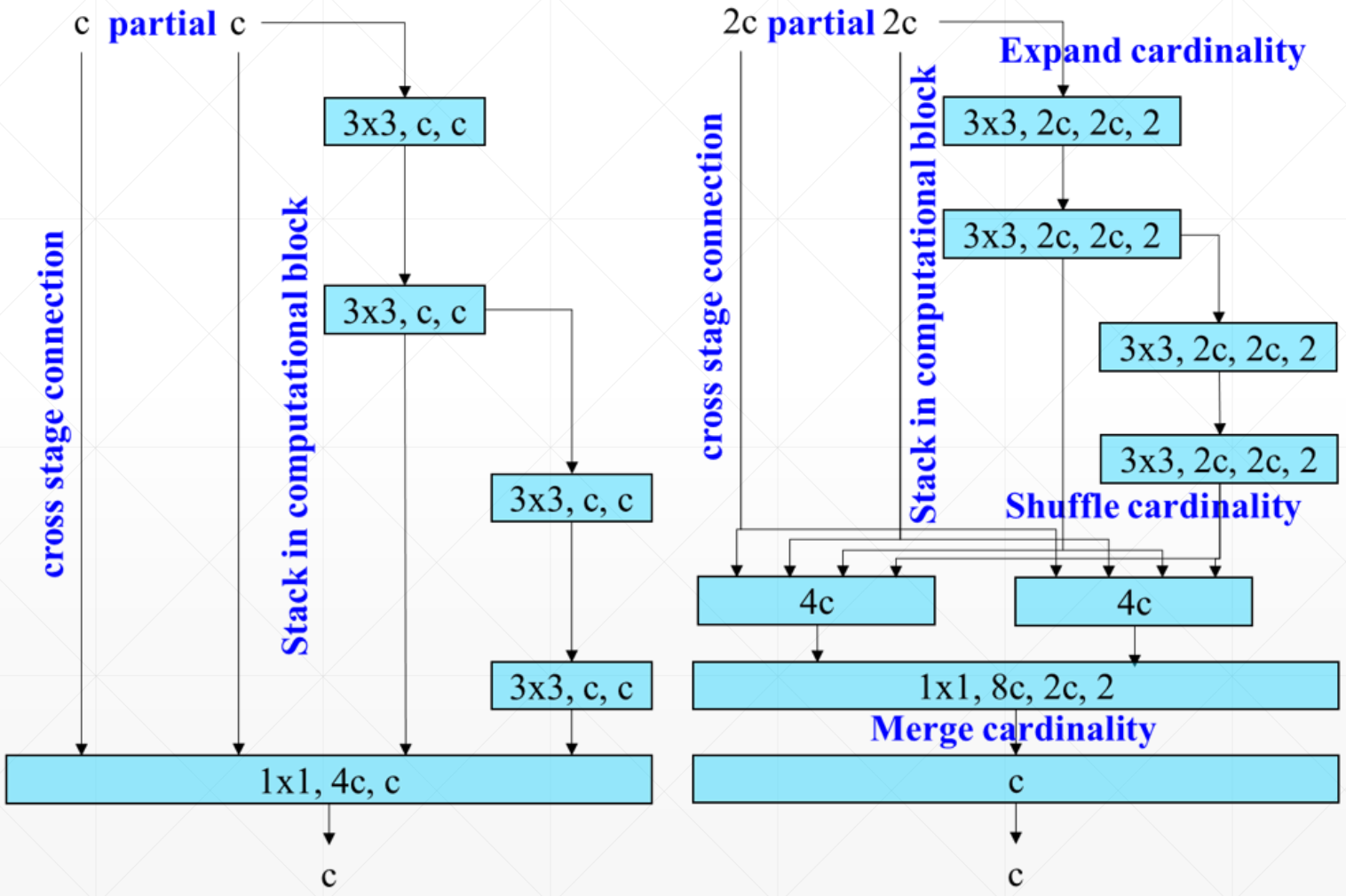


CSPVoVNet **ELAN**

Model	FLOPs	#Params	AP^{box}	AP^{mask}
YOLOR-v3 [34]	194.6G	64.3M	49.5%	40.9%
YOLOR-PRN	194.6G	64.3M	50.0%	41.0%
YOLOR-CSP	159.0G	54.3M	51.0%	41.1%
YOLOR-ELAN	**143.2G**	**34.5M**	**51.4%**	**41.5%**

Extended efficient layer aggregation networks

- Scaling up ELAN without modifying gradient path topology.

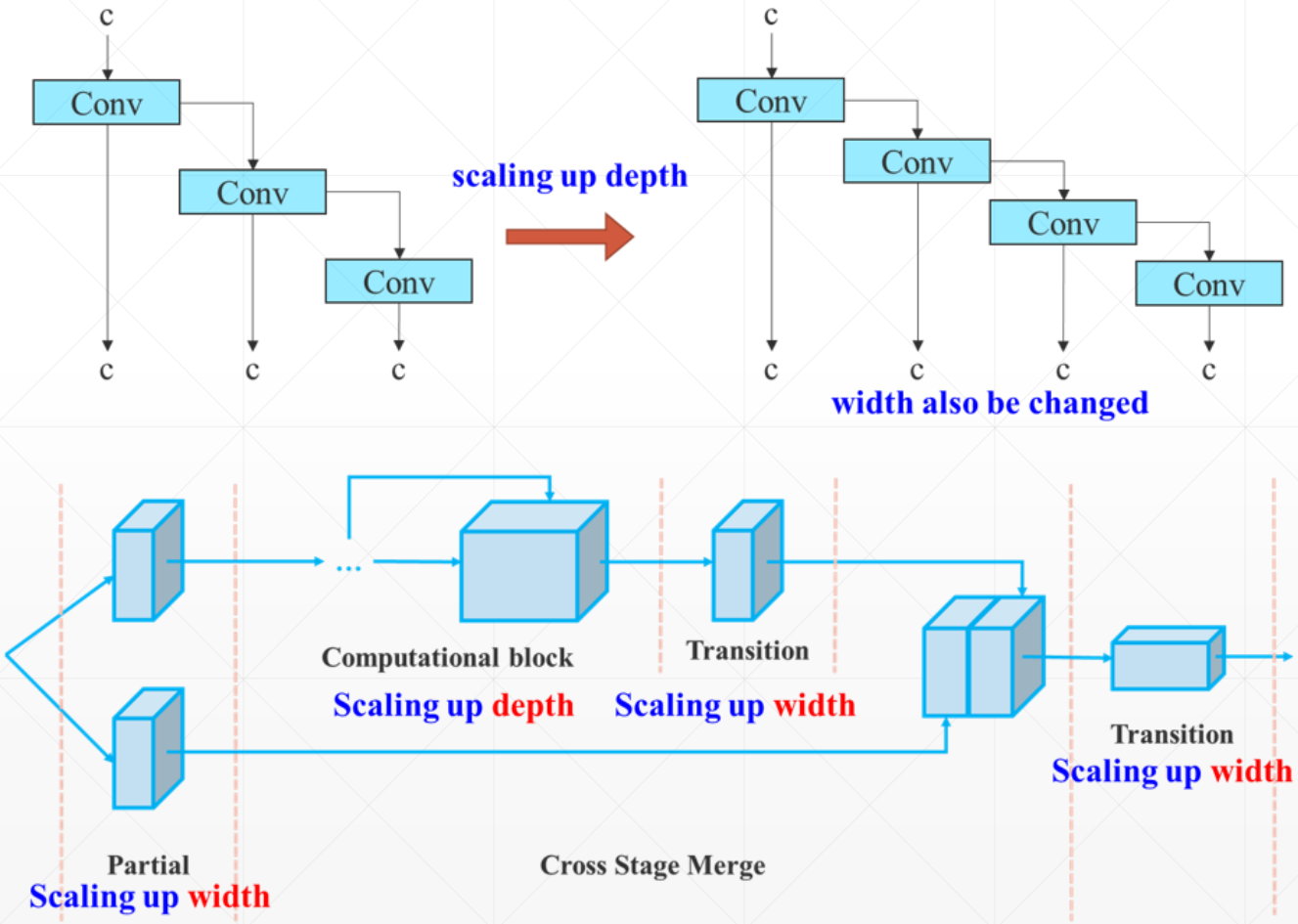


ELAN Extended ELAN

Model scaling for concatenation-based models

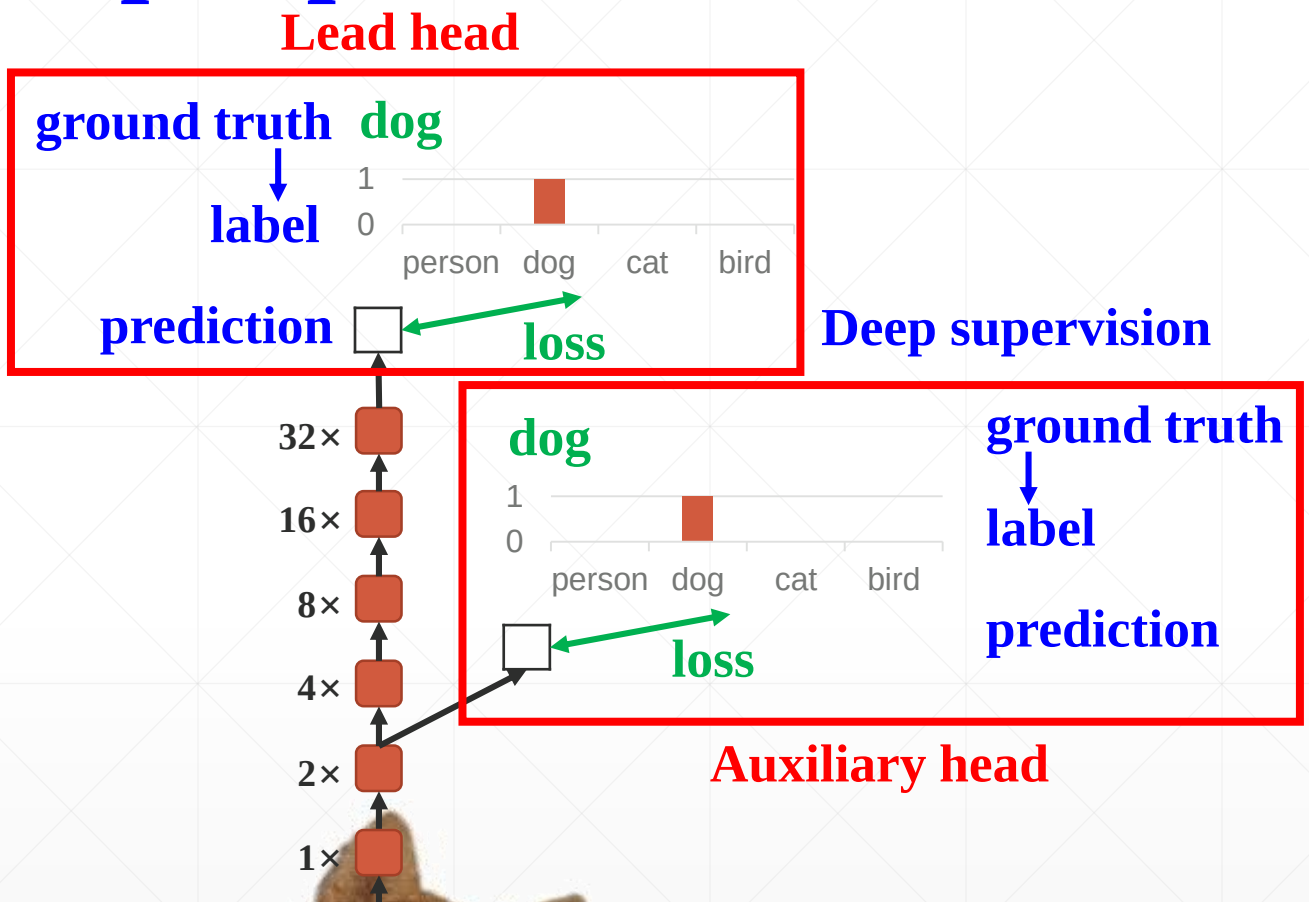
Model scaling for concatenation-based models

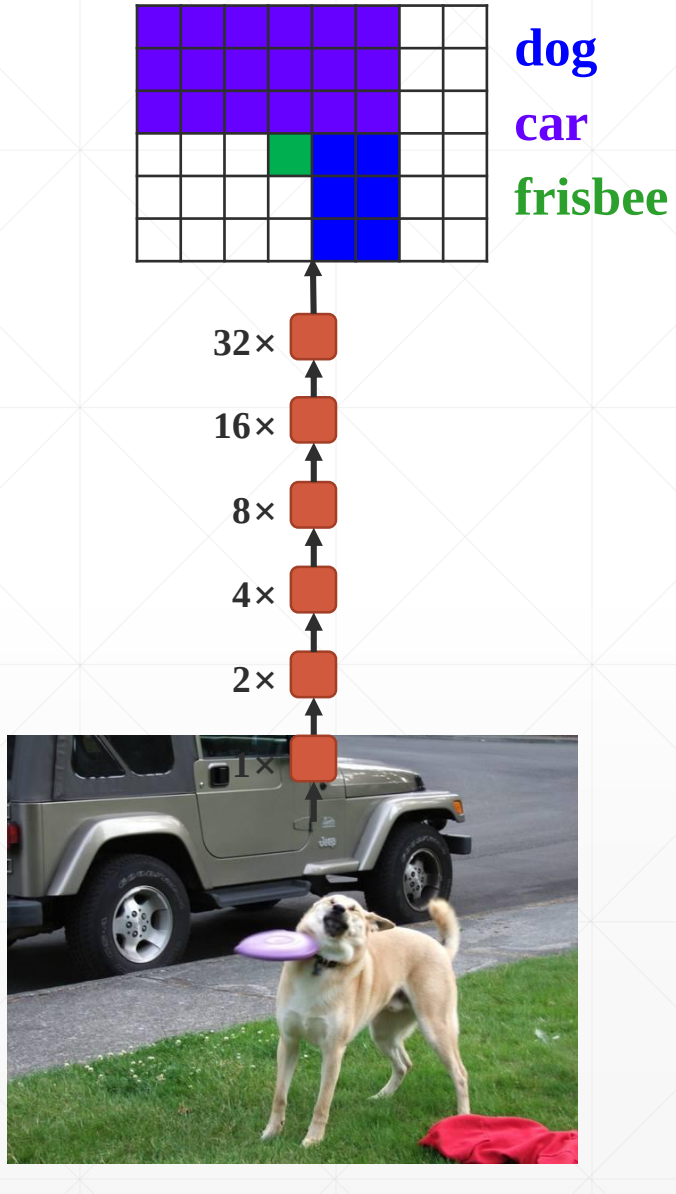
- We proposed to compound scaling width of transition layers and depth of computational blocks.



Dynamic label assignment meets deep supervision

Deep supervision & Label assignment

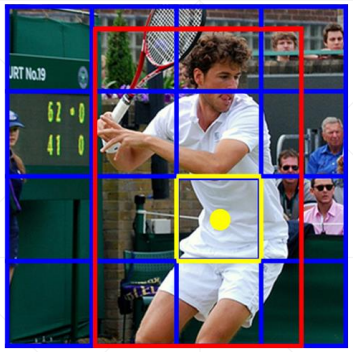




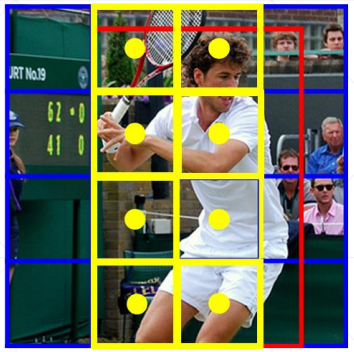
Detection (local prediction)

Dynamic label assignment (DLA)

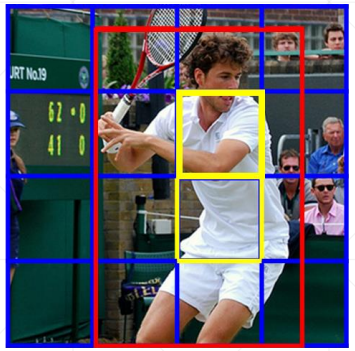
- **Traditional label assignment: ground truth + rules = label**



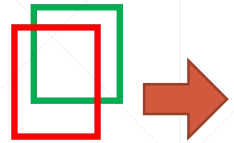
Object center

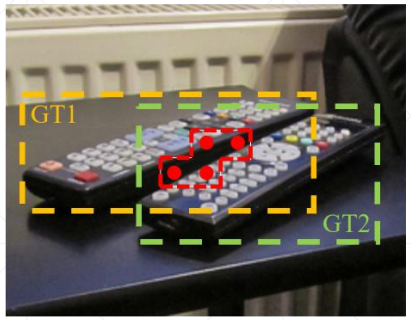


Grid center



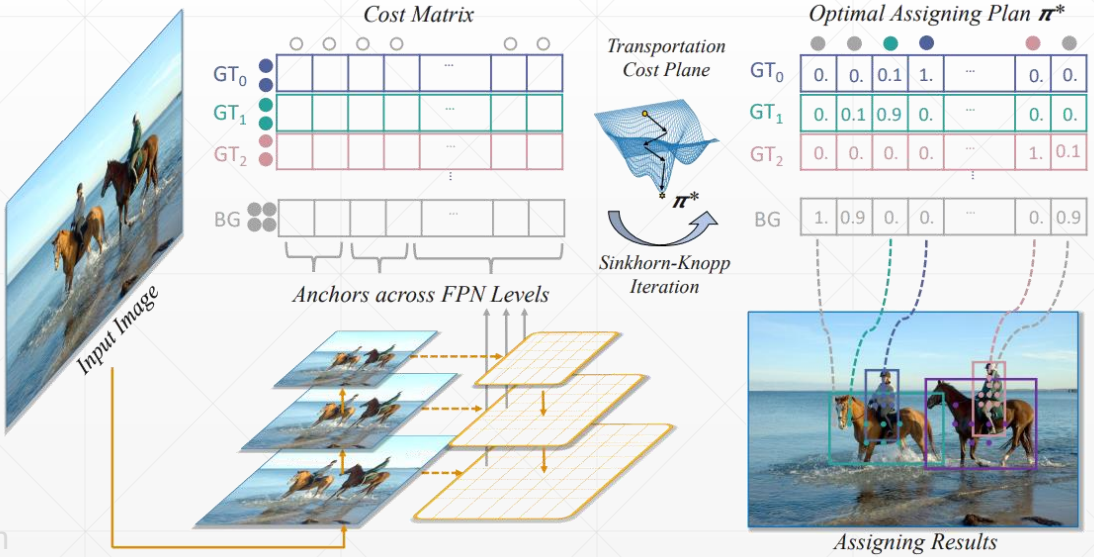
IoU threshold





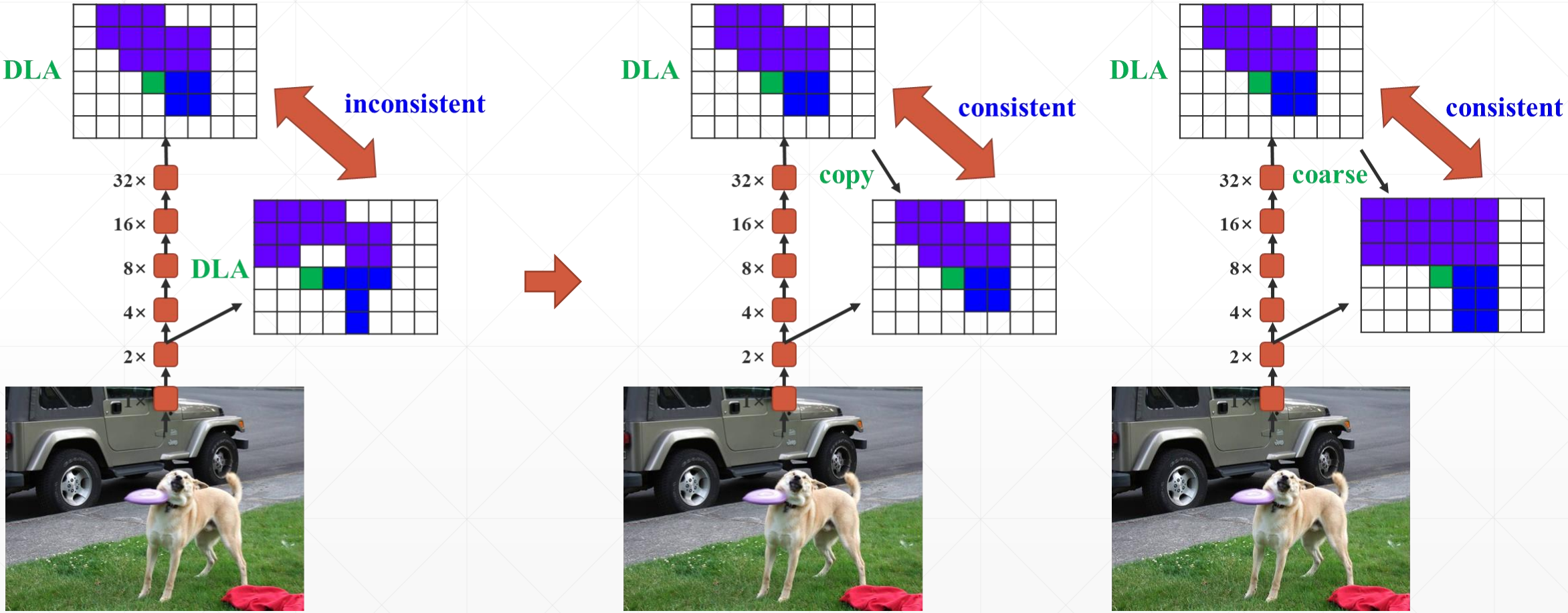
GT BBoxes A Set of Ambiguous Anchor Points

- **Dynamic label assignment: ground truth + optimization = label**



OTA: Optimal Transport Assignment for Object Detection
https://medium.com/@_Xing_Chen_/yolox-%E8%A9%B3%E7%B4%B0%E8%A7%A3%E8%AE%80-dd9bcbcff208

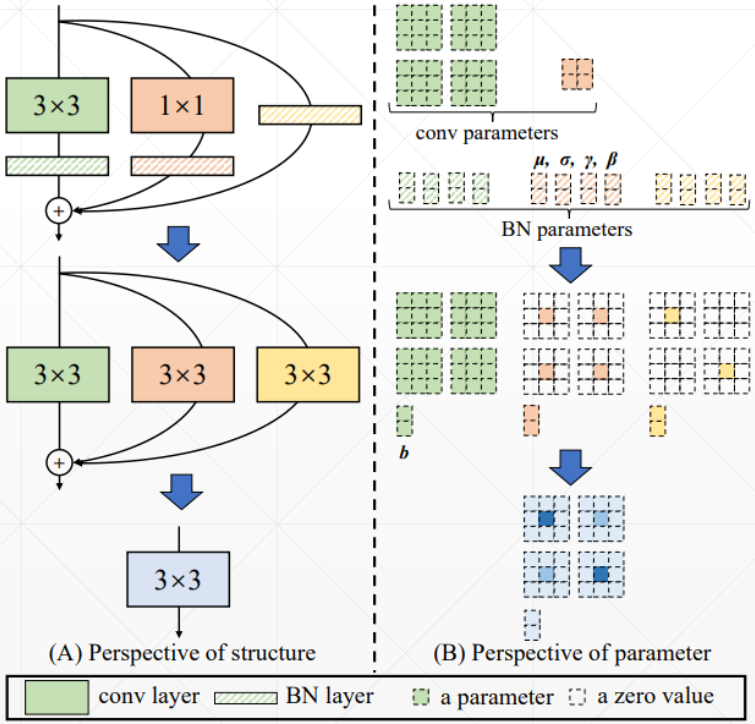
Deep supervision meets DLA

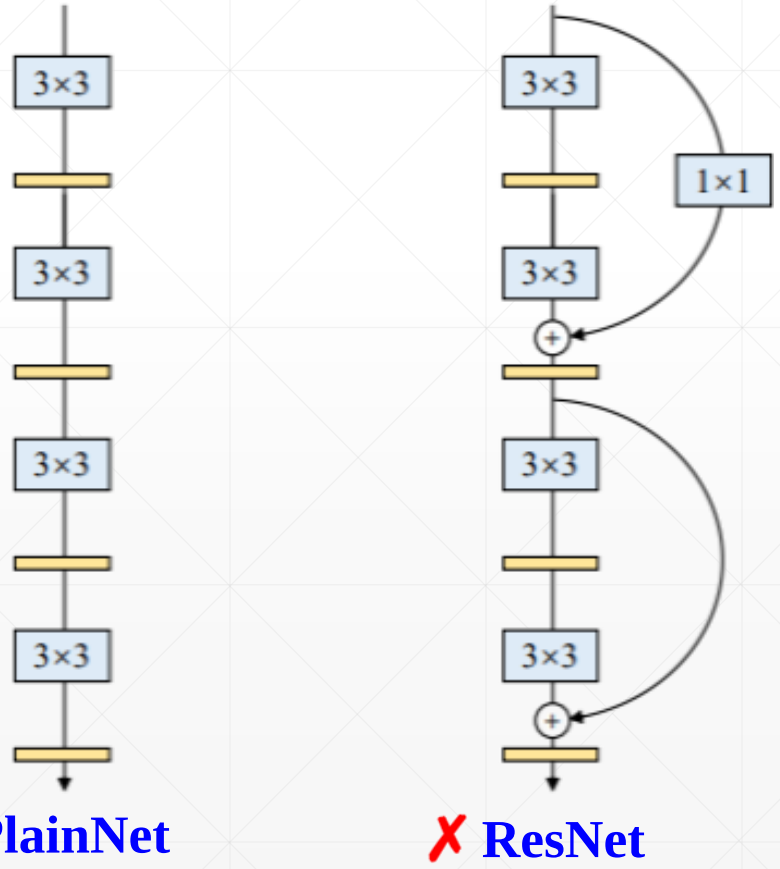


Re-parameterization model meets modern networks

RepConv

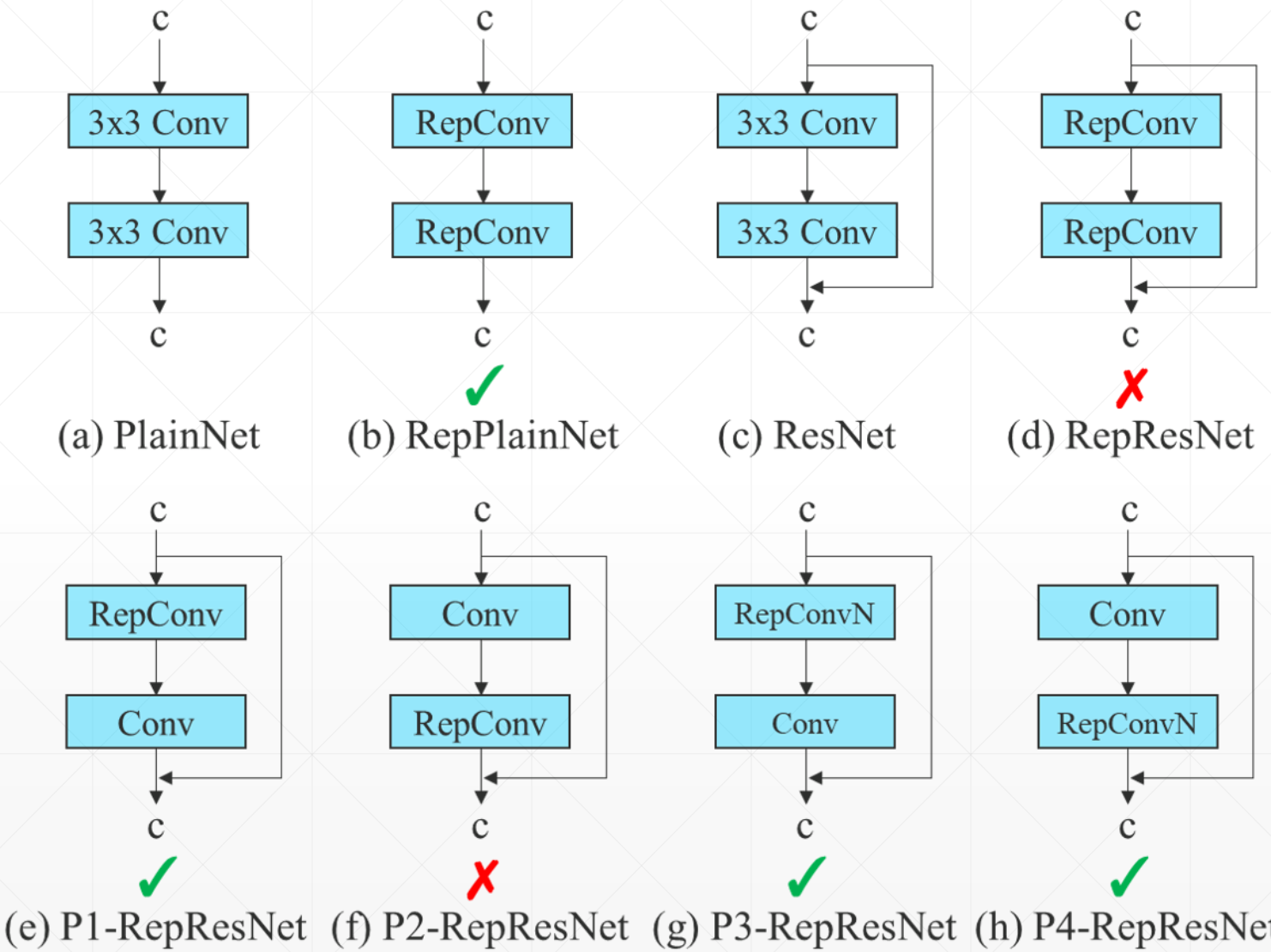
- RepConv can make very deep PlainNet converge.
- But it will make accuracy of modern networks drop, such as ResNet and DenseNet.





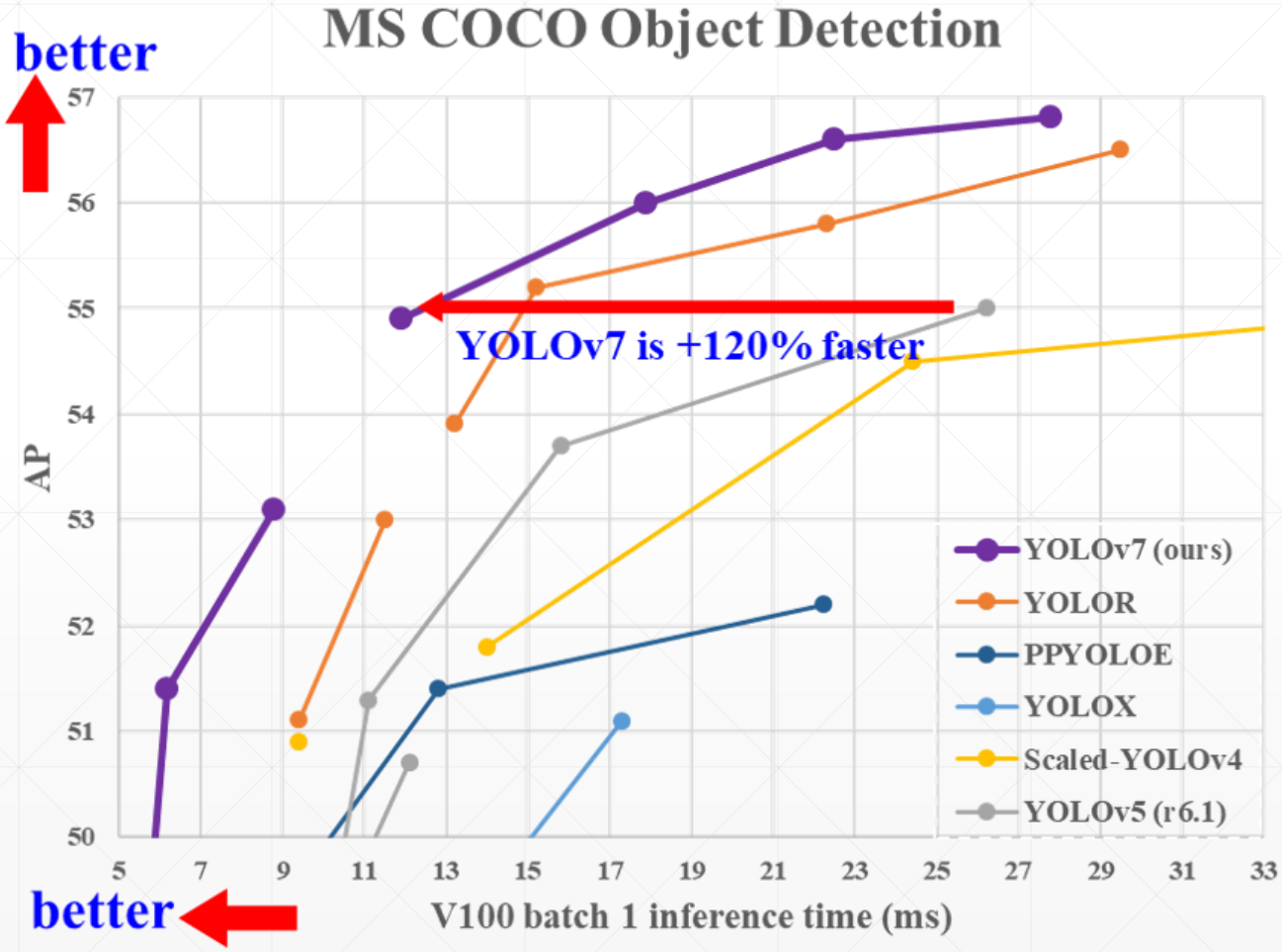
✓ PlainNet ✗ ResNet

Planned Re-parameterization model



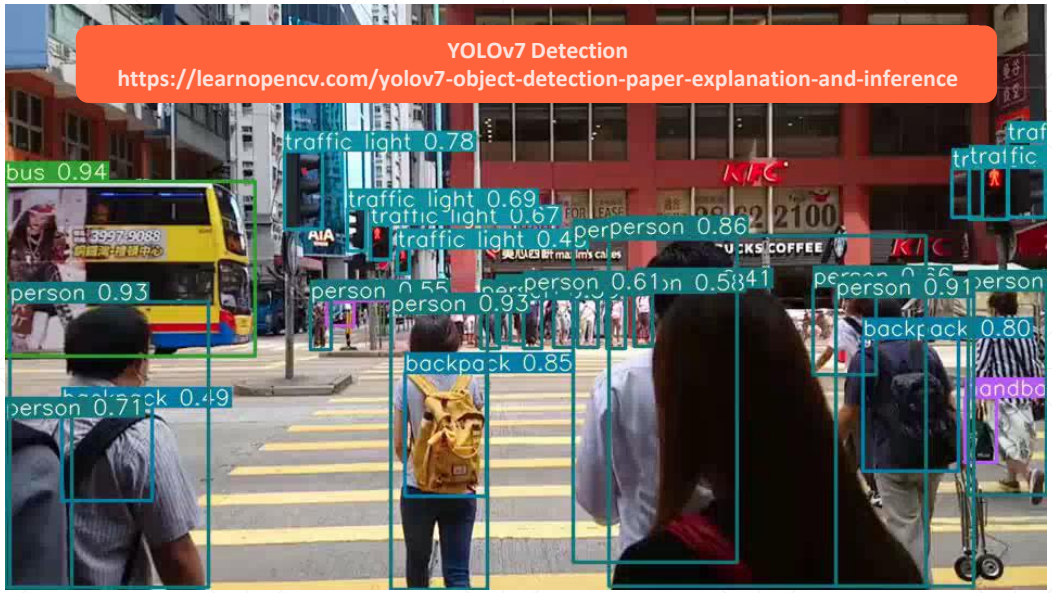
Results

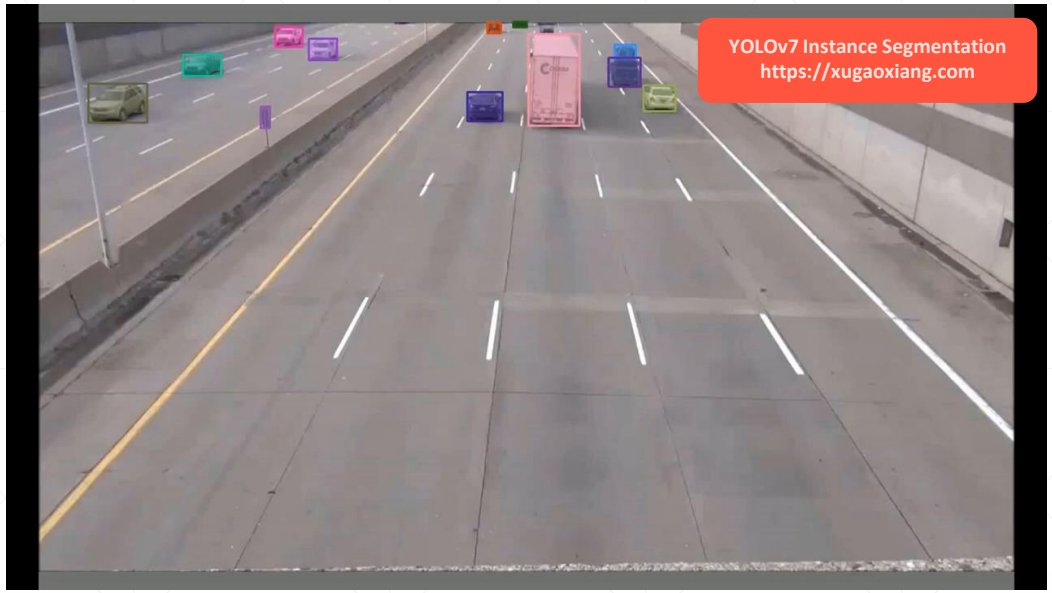
Results


MS COCO Object Detection
better
57
56
55
54
53
52
51
50
AP
YOLOv7 is +120% faster
YOLOv7 (ours)
YOLOR
PPYOLOE
YOLOX
Scaled-YOLOv4
YOLOv5 (r6.1)
5
7
9
11
13
15
17
19
21
23
25
27
29
31
33
better
V100 batch 1 inference time (ms)

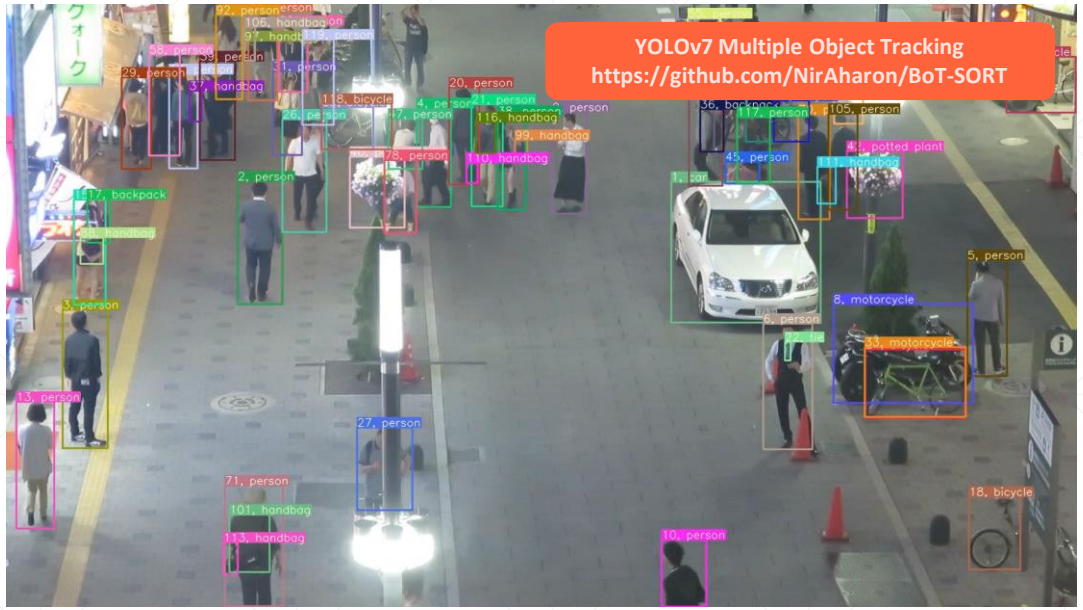
- Motivation & Introduction
- Proposed Method & Results
- **Applications & Conclusion**

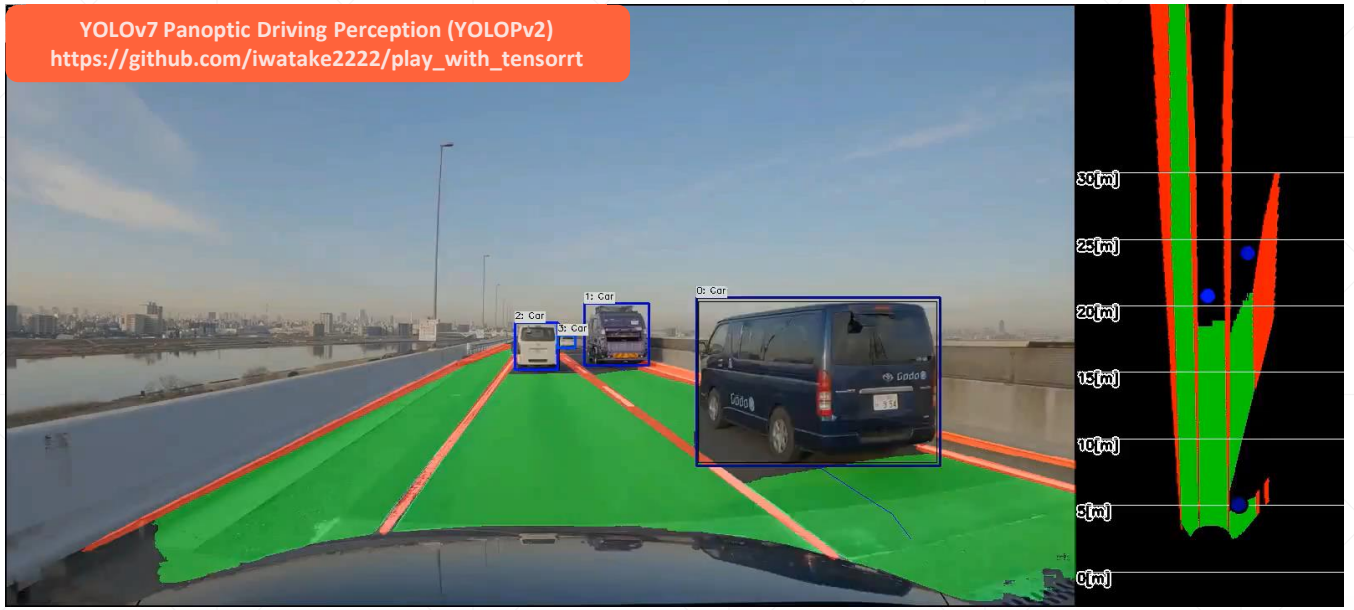
Applications (1/2)

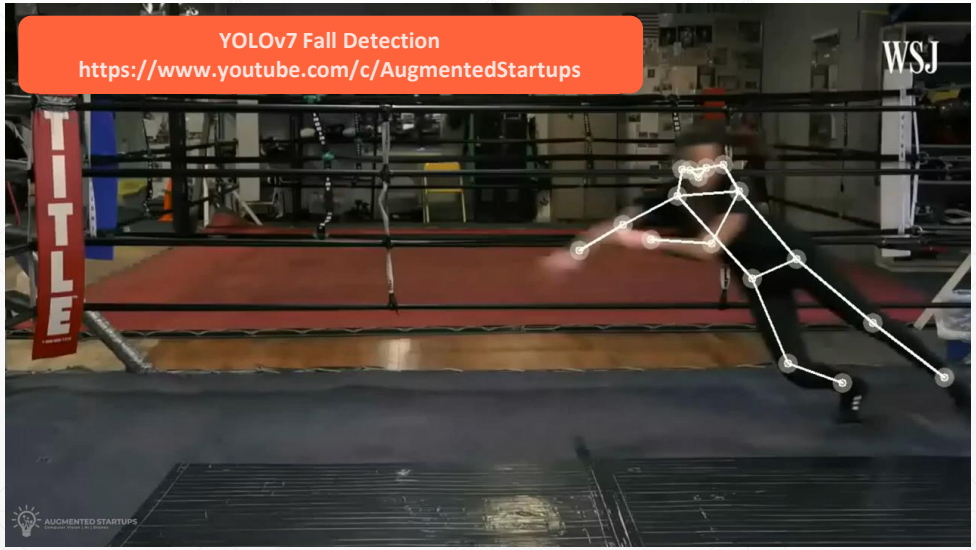


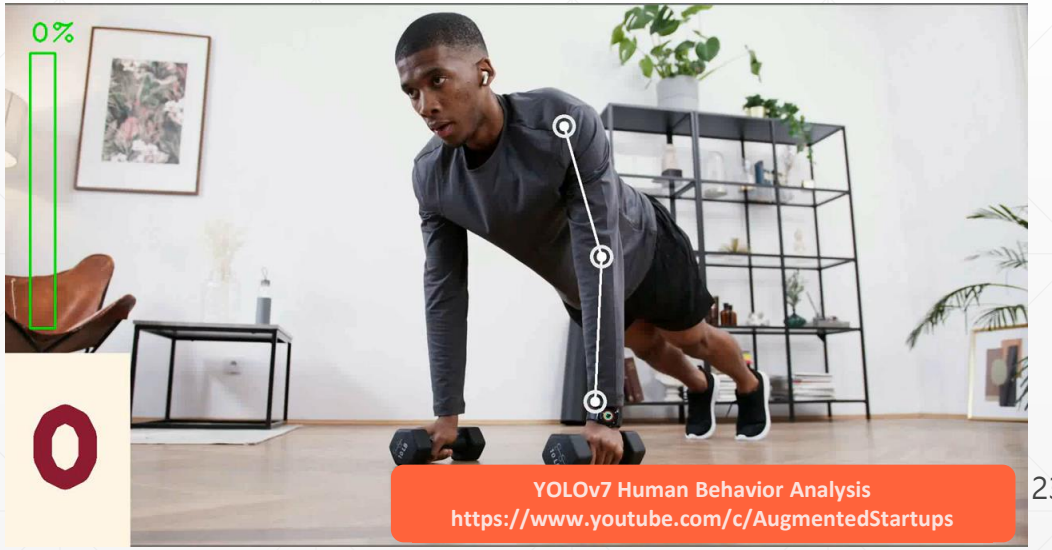


Applications (2/2)









Conclusions

- In this paper, we
  - 1. propose a new architecture of real-time object detector and the corresponding model scaling method.
  - 2. find that the evolving process of object detection methods generates new research topics.
  - 3. solve the replacement problem of re-parameterization module and the allocation problem of dynamic label assignment.
  - 4. propose the trainable bag-of-freebies method to enhance the accuracy of object detection.

Thanks For Listening

Q&A