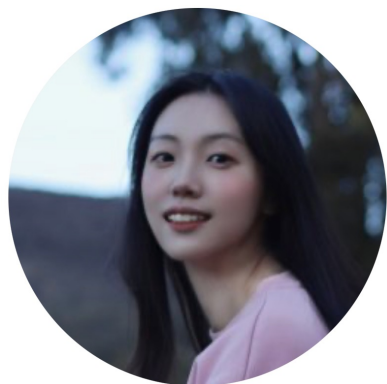


BoS: Bounding Box Stability against Feature Dropout Reflects Detector Generalization across Environments



Yang Yang



Wenhai Wang



Zhe Chen



Jifeng Dai



Liang Zheng



上海人工智能实验室
Shanghai Artificial Intelligence Laboratory



ICLR 2024

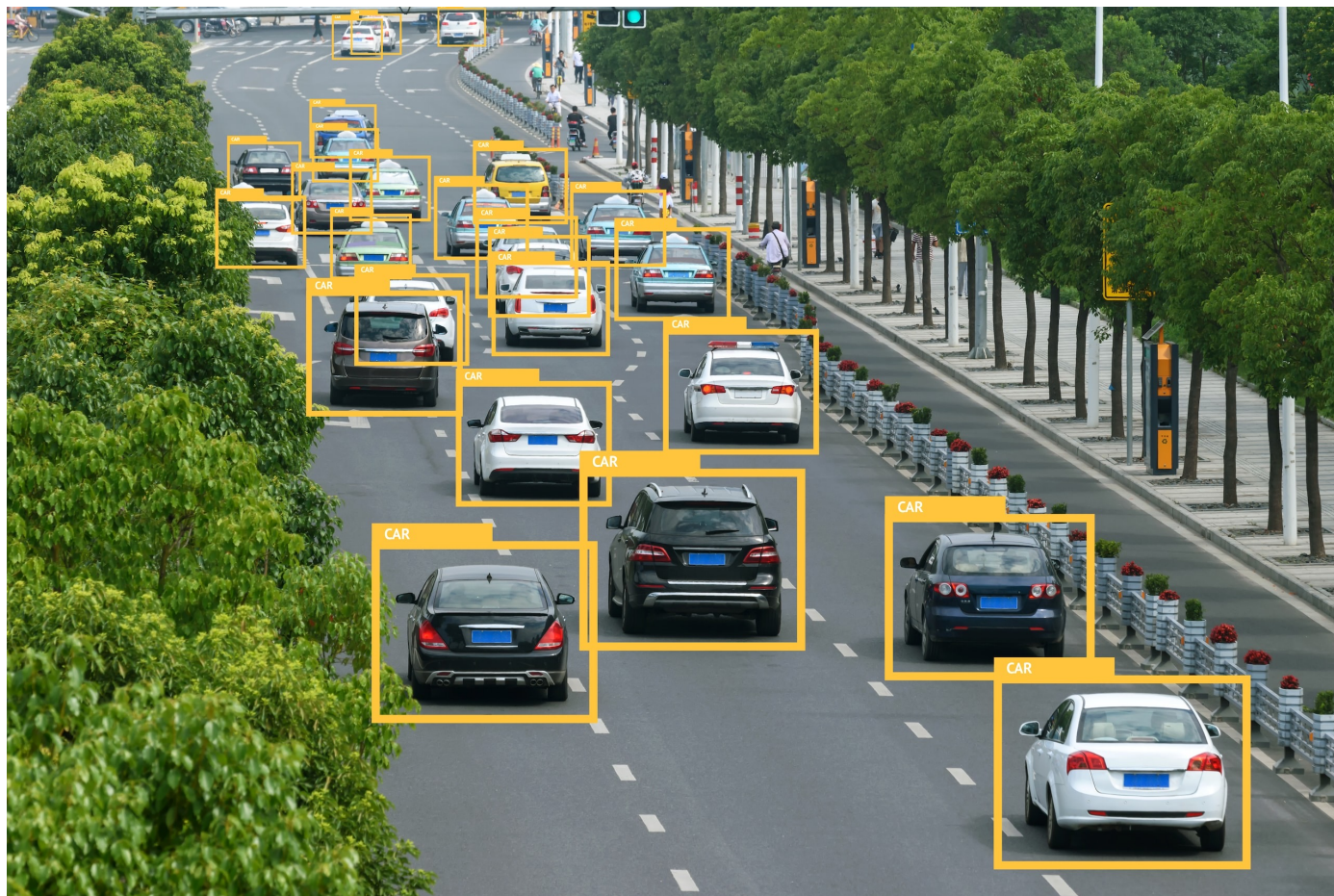
Notable-Top-5% (Spotlight)

<https://github.com/YangYangGirl/BoS>

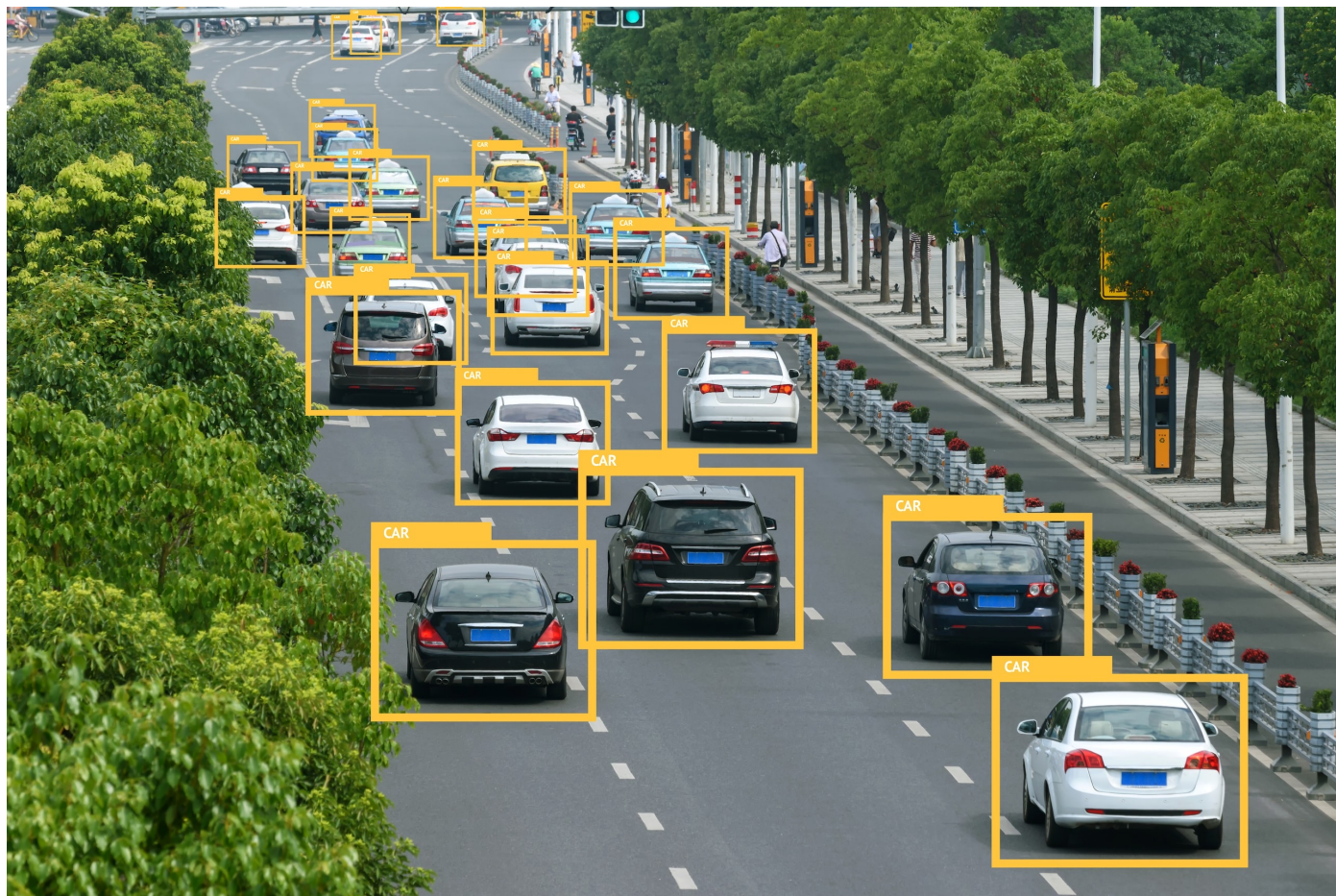


ICLR
International Conference On
Learning Representations

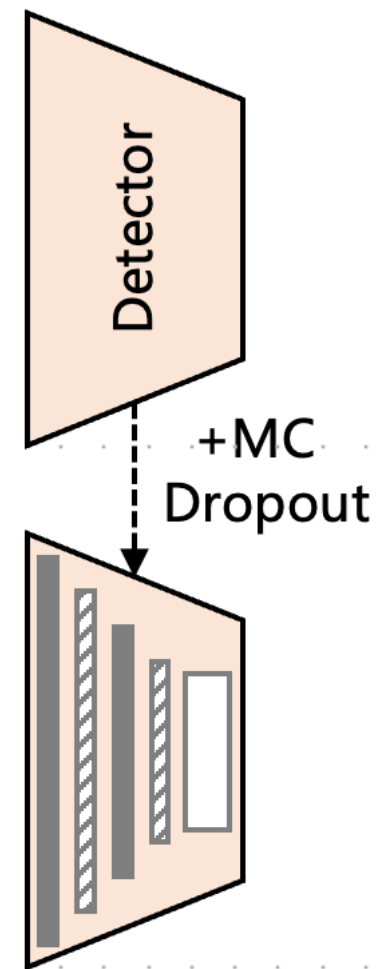
Object Detection



Object Detection



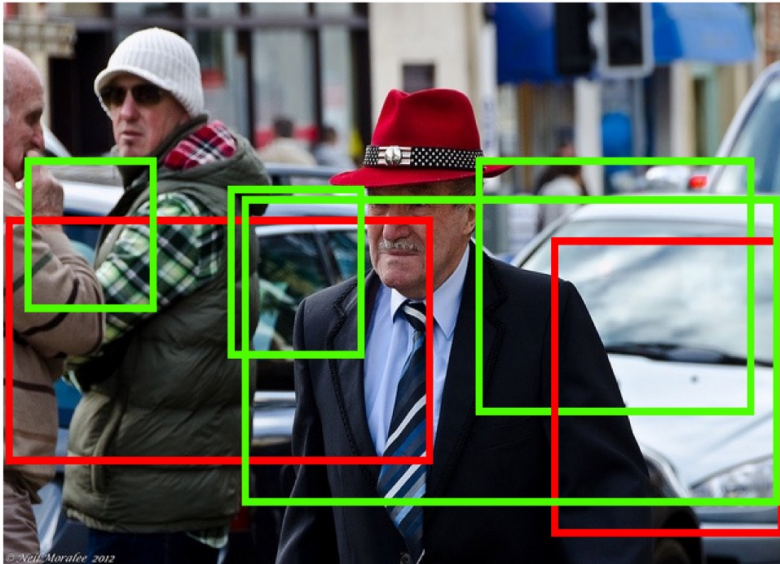
Motivation 



Motivation

 Original Detection

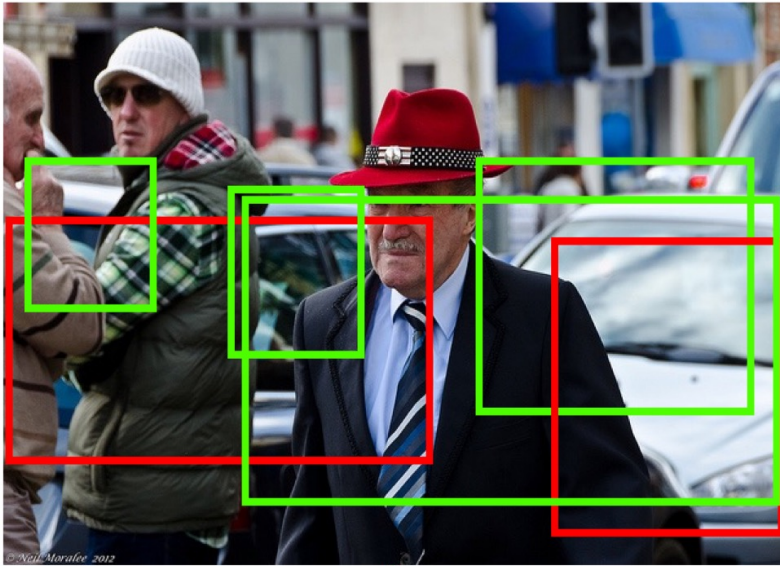
(a) Unstable Predictions



Motivation

 Original Detection

(a) Unstable Predictions



 Perturbed Detection

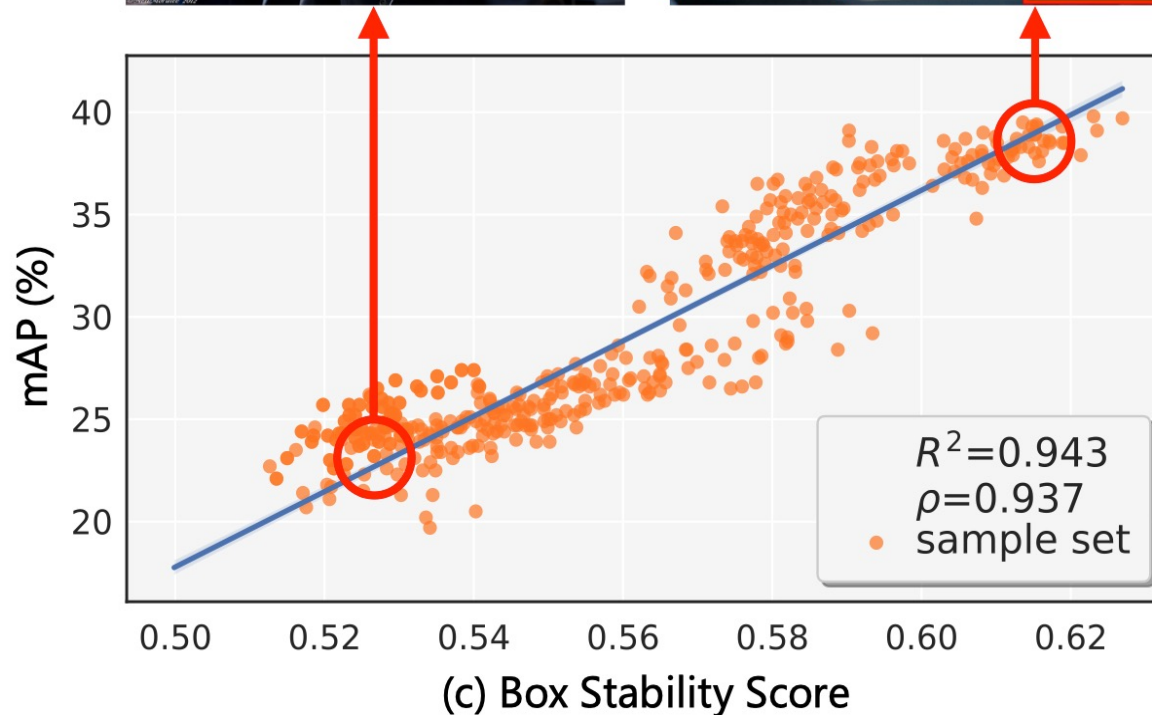
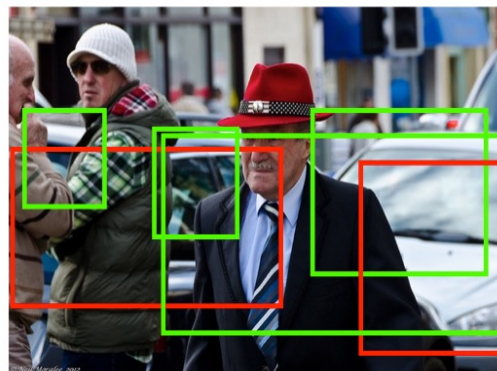
(b) Stable Predictions



Our Findings

Strong correlation between bounding box stability and detection accuracy under various real-world test environments.

 Original Detection  Perturbed Detection
(a) Unstable Predictions (b) Stable Predictions



Our Findings

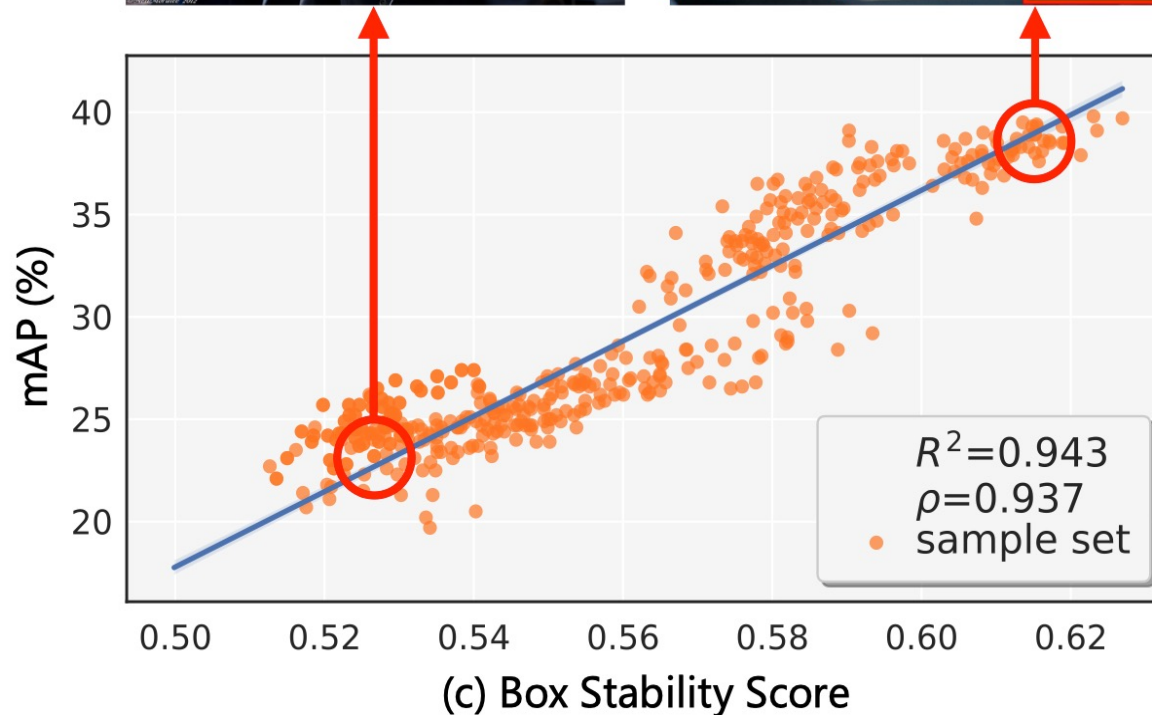
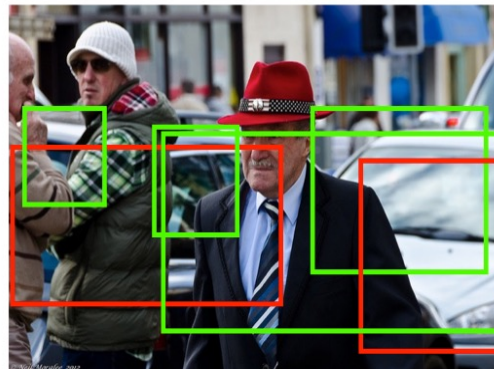
Strong correlation between bounding box stability and detection accuracy under various real-world test environments.

Use Case ??

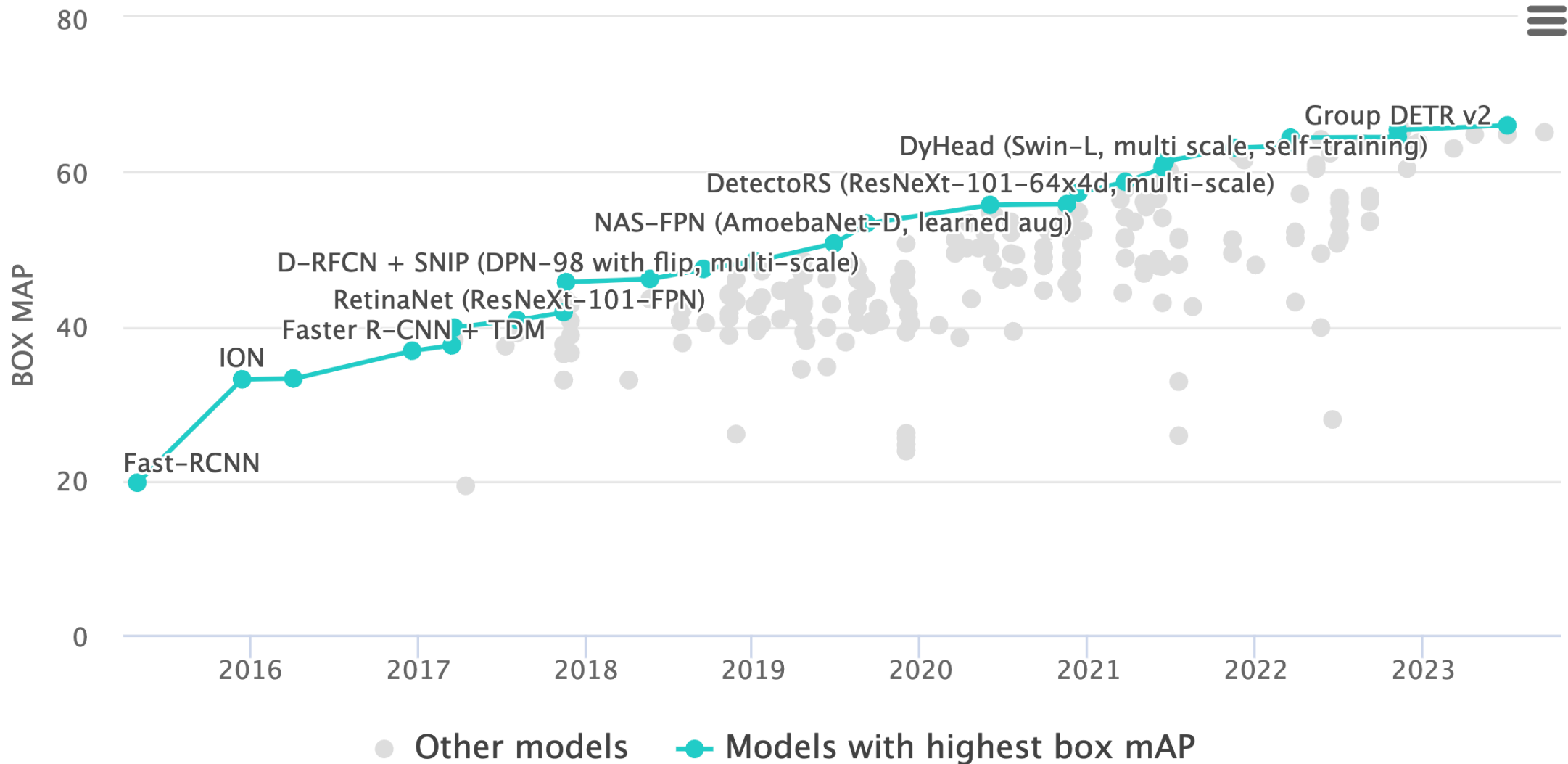
 Original Detection  Perturbed Detection

(a) Unstable Predictions

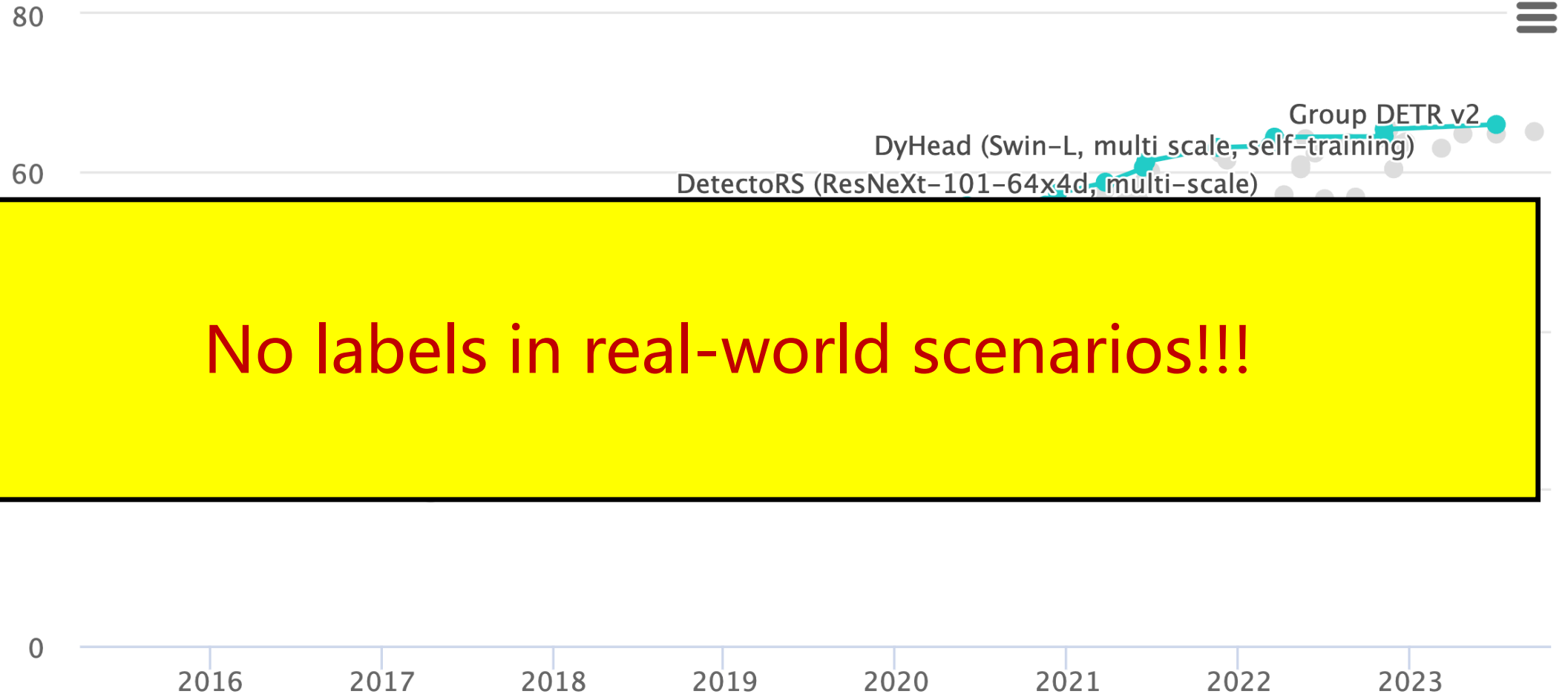
(b) Stable Predictions



Object Detection on COCO test-dev Benchmarks



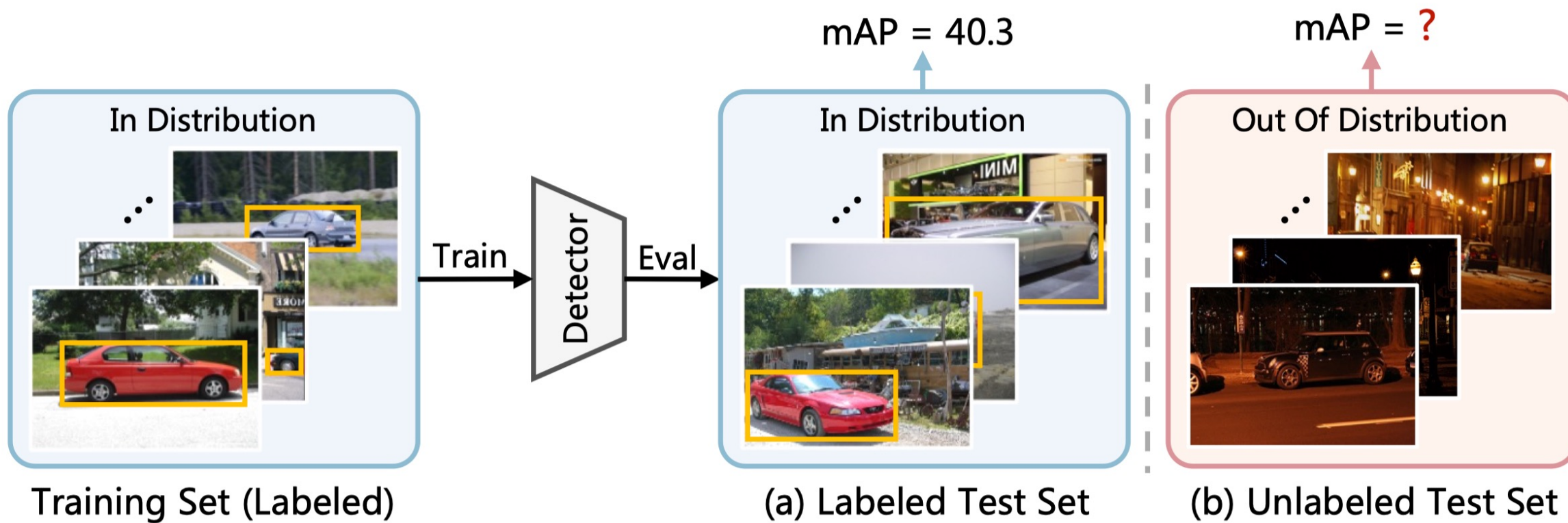
Object Detection on COCO test-dev Benchmarks



No labels in real-world scenarios!!!

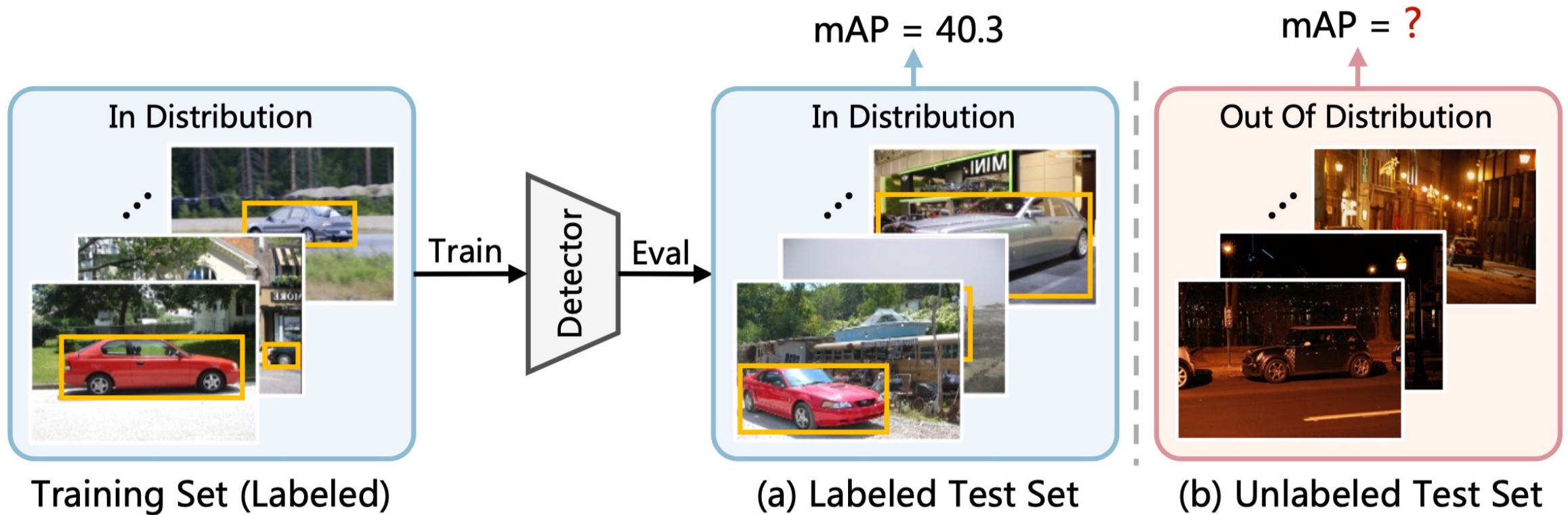
● Other models ● Models with highest box mAP

Use Case: Label-free Detector Evaluation



$$\text{mAP} = \omega_1 * ? + \omega_0$$

Use Case: Label-free Detector Evaluation



$$mAP_{\text{?}} = \omega_1 * ? + \omega_0$$

a brand-new problem

Our Findings

- Strong correlation between bounding box stability and detection accuracy under various real-world test environments.

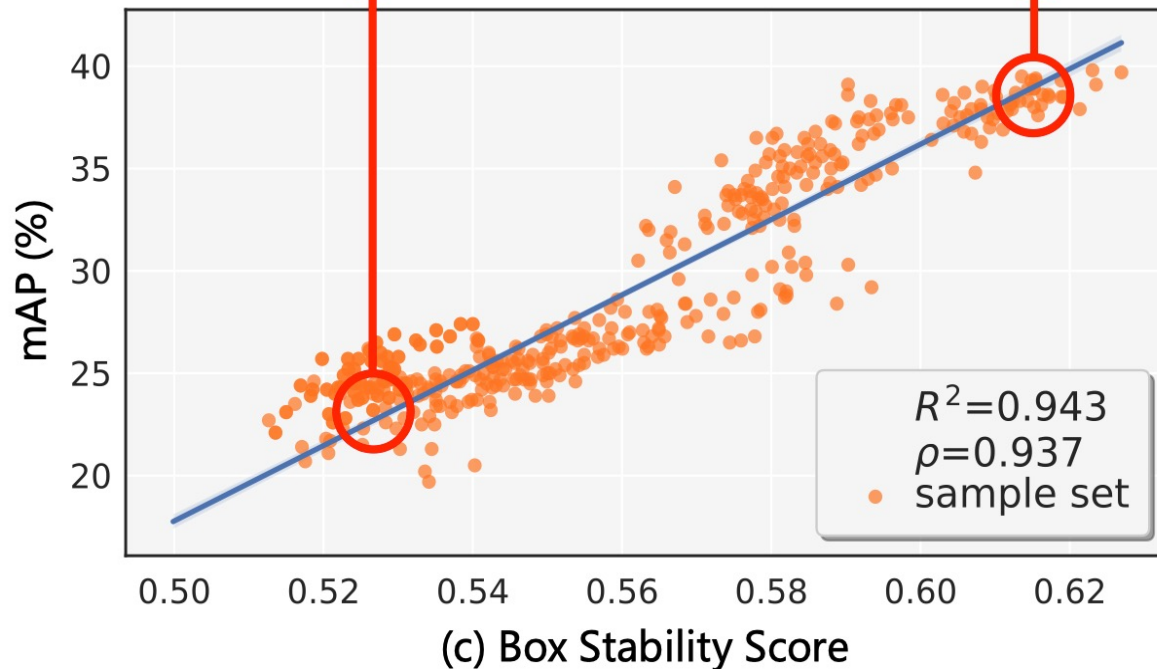
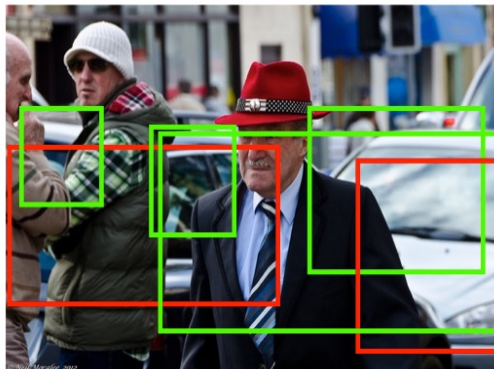
$$mAP = \omega_1 BoS + \omega_0$$

 Original Detection

 Perturbed Detection

(a) Unstable Predictions

(b) Stable Predictions



Establish Meta Dataset

Seed Set



Sample Set 1



Sample Set 2



Sample Set 3



Sample Set 4



Sample Set 5



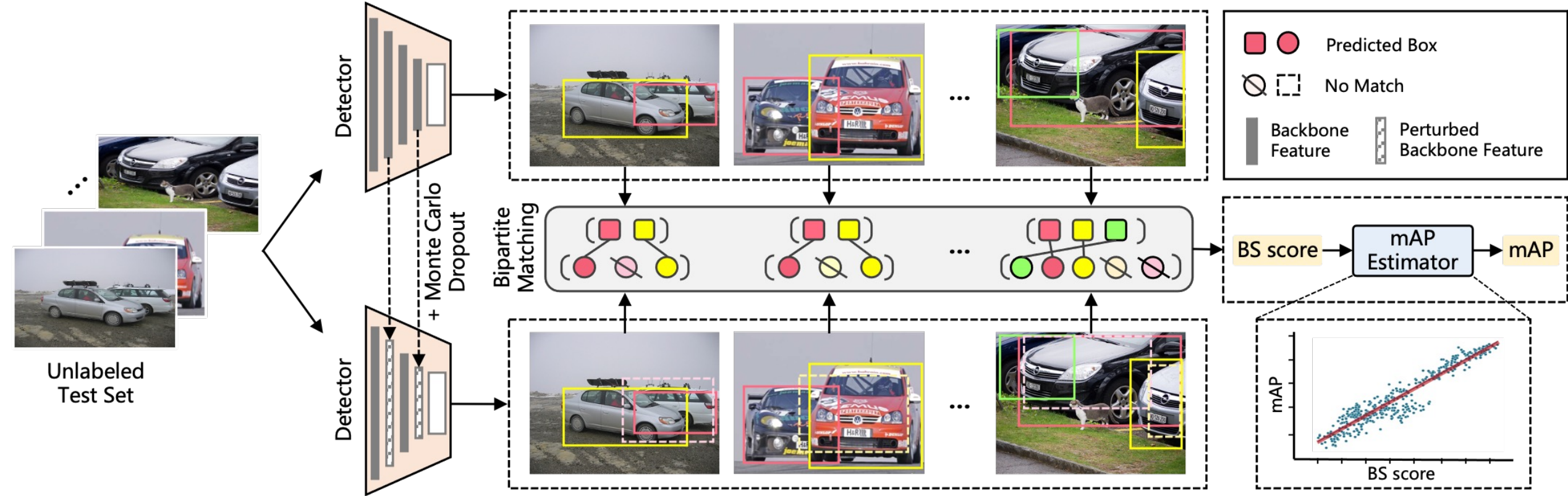
Sample Set 6



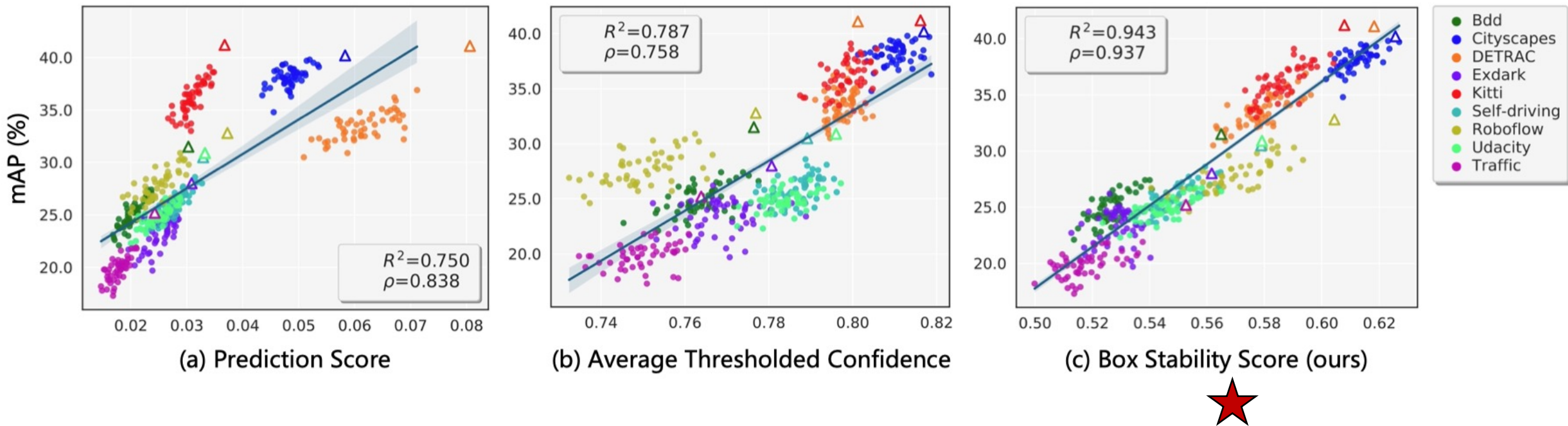
Sample Set 7



A workflow summary of **Box Stability Score (BoS)**



Correlation between different measurements and mAP



" Δ " of different colors represents nine real-world datasets

Each point " \bullet " of different colors represents a sample set generated from different seed sets

Main Experiment

Method	COCO 34.2	BDD 31.5	Cityscapes 40.2	DETRAC 41.1	Exdark 28.0	Kitti 41.2	Self-driving 30.5	Roboflow 32.8	Udacity 30.9	Traffic 25.2	Avg. RMSE ↓
Prediction Score [22]	8.14	3.22	5.50	15.83	0.81	11.80	1.26	2.48	1.54	1.99	5.26
Entropy Score [41]	7.06	1.37	8.40	16.20	0.14	13.08	0.93	3.14	1.23	3.6	5.52
AC [19]	6.67	3.32	9.60	30.95	1.26	13.93	2.31	1.77	2.92	3.3	7.60
ATC [16]	10.35	3.20	5.83	8.21	1.44	6.34	0.91	5.13	1.94	1.23	4.46
FD [10]	9.17	2.78	13.03	12.29	5.94	14.80	2.22	4.48	1.32	4.14	7.02
box stability score (ours)	1.26±0.32	1.90±0.06	0.89±0.28	1.84±0.05	1.47±0.19	4.38±0.07	1.92±0.09	4.79±0.10	1.61±0.06	2.43±0.09	2.25±0.13
box stability score + Prediction	2.11±0.29	2.03±0.06	1.13±0.26	0.04±0.02	1.42±0.18	5.56±0.06	1.71±0.08	4.69±0.10	1.42±0.05	2.24±0.09	2.24±0.12

Method comparison of mAP estimation for vehicle detection

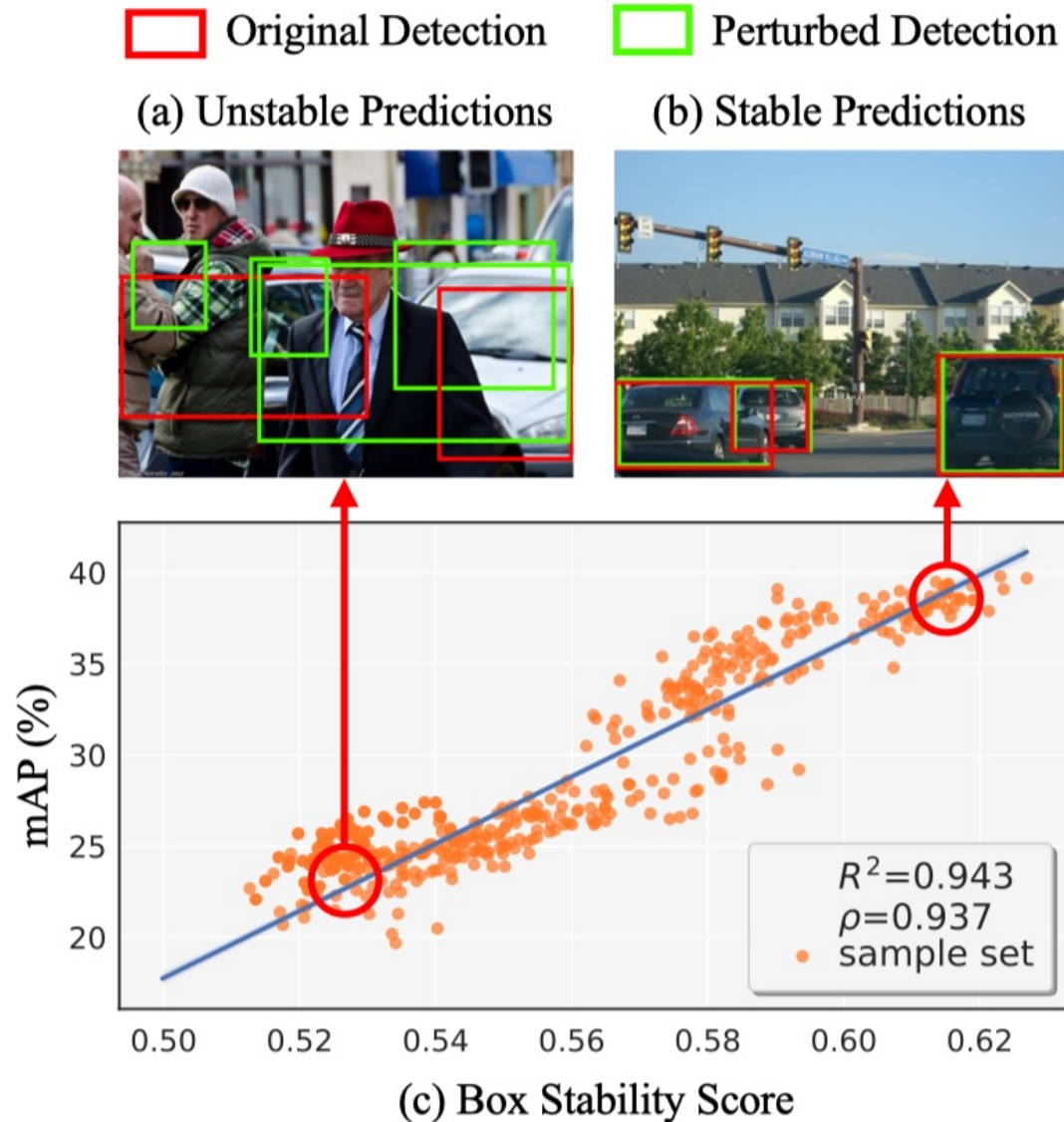
Method	COCO 26.7	Caltech 16.2	CrowdHuman 33.5	Cityscapes 19.0	Self-driving 16.4	Exdark 21.6	EuroCity 16.8	Kitti 13.3	CityPersons 11.8	Avg. RMSE ↓
Prediction Score [22]	8.49	3.68	6.55	0.76	3.20	5.58	0.86	0.99	6.28	4.04
Entropy Score [41]	5.91	1.02	9.44	1.06	1.63	3.64	0.13	2.23	5.71	3.42
AC [19] or ATC [16]	5.25	1.38	17.25	0.79	1.33	3.18	0.39	1.69	5.99	4.13
FD [10]	3.88	3.62	3.29	5.19	3.41	3.37	3.93	2.98	4.21	3.76
box stability score (ours)	2.26±0.05	2.53±0.07	6.39±0.02	0.18±0.11	4.42±0.06	0.74±0.07	1.72±0.04	5.16±0.08	6.26±0.15	3.29±0.07
box stability score + Prediction	7.04±0.05	2.18±0.07	4.63±0.02	0.43±0.11	3.29±0.06	0.58±0.07	2.33±0.04	3.99±0.08	6.88±0.15	3.48±0.07

Method comparison of mAP estimation for pedestrian detection

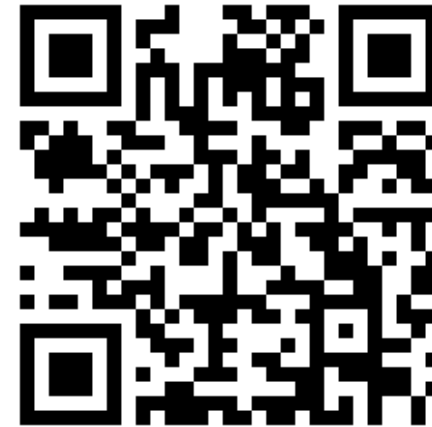
Main Contributions

- We report a strong positive correlation between the detection mAP and BS score.
- We demonstrate an interesting use case of this strong correlation: estimating detection accuracy in terms of mAP without test ground truths. To our best knowledge, we are the first to propose the problem of unsupervised evaluation of object detection.
- To show the effectiveness of the BS score in estimating detection mAP, we conduct experiments on two detection tasks: vehicle detection and pedestrian detection.

BoS: Bounding Box Stability against Feature Dropout Reflects Detector Generalization across Environments



Thank You!



Project Page



Code