Semantic View in Autonomous Driving using Waymo Dataset

Contributors :周禮宏, 唐郁秀,紀昕妤, 賴煜翔和顏振宇 (*sorted by strokes of last name*)

The Success of Self-driving Vehicle

- **Understanding the environment**: commonly a 3D semantic HD map at the back-end precisely recorded the environment
- **Self-location**: an on-the-fly self-localization system puts the vehicles accurately inside the 3D world, so that it can plot a path to every target location
- **Semantics in the view:** a 3D perceptual system detects other moving objects, guidance signs and obstacles on the road, in order to avoid collisions and perform correct actions.

Wang et al.(2019)

Public Autonomous Driving Datasets

KIT
Karlsruher Institut für Technologie

HRI
Honda Research Institute US

WAYMO
Open
Dataset

CANBUS
NUSCENES

ARGO AI

apolloscape

KAIST

Agency for Science, Technology and Research
SINGAPORE
SUTD
SINGAPORE UNIVERSITY OF TECHNOLOGY AND DESIGN
CentraleSupélec
UNIVERSITY OF SURREY
ROYAL ACADEMY OF ENGINEERING

A*3D DATASET: TOWARDS AUTONOMOUS DRIVING IN CHALLENGING ENVIRONMENT

ORI

Pitropov et al. (2020)

WAYMO Open Dataset

The field of machine learning is changing rapidly. Waymo is in a unique position to contribute to the research community with one of the largest and most diverse autonomous driving datasets ever released.

Access Waymo Open Dataset

Sign in with Google

Waymo Open Automated Driving Dataset

- **large scale, high quality, diverse dataset**
  - from 1,950 driving segments that each span 20 seconds, corresponding to 200,000 frames at 10 Hz per sensor
  - well synchronized and calibrated high quality **LiDAR and five front-and-side-facing camera data**
  - captured across a range **of urban and suburban geographies**
  - labels for four object classes: vehicles, pedestrians, cyclists, and signs

- **strong baselines for 2D as well as 3D detection and tracking tasks**
  - Data have been **annotated with 2D (camera image) and 3D (LiDAR) bounding boxes**.

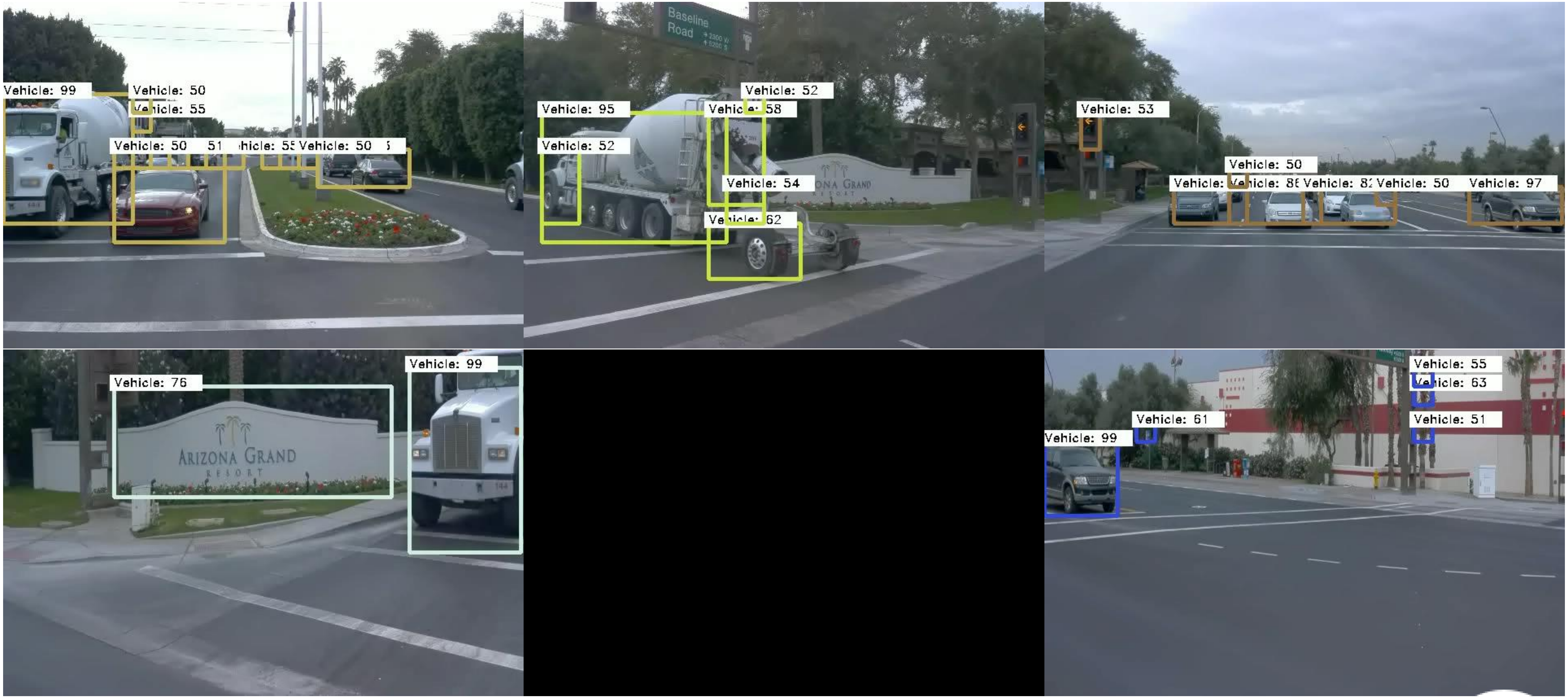
Sun et al. (2019)

Waymo Open Dataset. Available online: https://waymo.com/open/ (accessed on 31 May 2020).

Machine Perception Task

- Overall:
  - Video (or Image) : image classification, **object detection**, object tracking (Perez et al., 2017), semantic segmentation as well as instance segmentation(Paszke et al., 2016)

  - LiDAR: distance-aware warning
- **Specifically for Waymo dataset :**
  - driving action prediction (Gu et al., 2020) and data augmentation (Via, 2020)


Vehicle: 99
Vehicle: 50
Vehicle: 55
Vehicle: 50
51
Vehicle: 50
Vehicle: 95
Vehicle: 52
Vehicle: 52
58
Vehicle: 54
62
Baseline Road
Vehicle: 53
Vehicle: 50
Vehicle: 50
Vehicle: 97
Vehicle: 76
Vehicle: 99
ARIZONA GRAND
RESORT
Vehicle: 55
Vehicle: 63
Vehicle: 61
Vehicle: 51
Vehicle: 99

Object Detection in 20 years

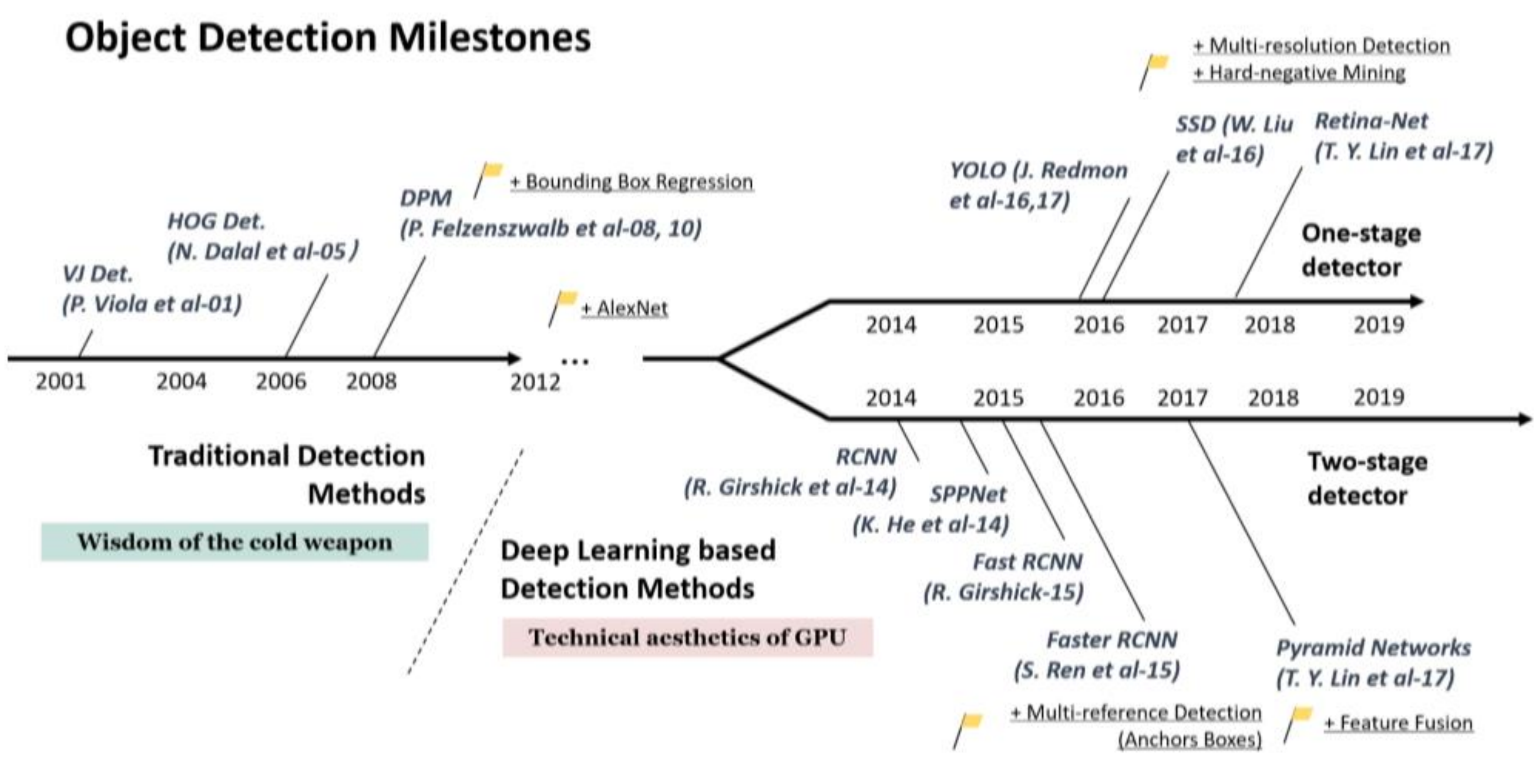


Fig. 2. A road map of object detection. Milestone detectors in this figure: VJ Det. [10, 11], HOG Det. [12], DPM [13–15], RCNN [16], SPPNet [17], Fast RCNN [18], Faster RCNN [19], YOLO [20], SSD [21], Pyramid Networks [22], Retina-Net [23].

Zou et al. (2019)

- Before 2014, traditional methods were to recognize the component.
- After 2014, learning based methods were to represent high-level feature with proposal detection and verification .

Objects Detection - Faster R-CNN

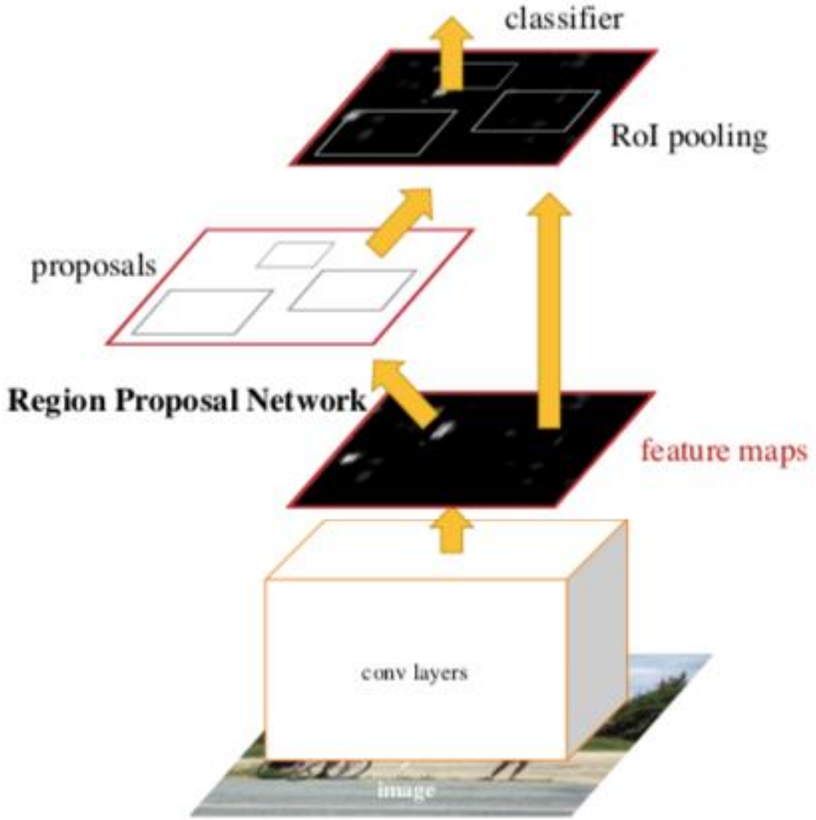


Figure 2: Faster R-CNN is a single, unified network for object detection. The RPN module serves as the 'attention' of this unified network.

- Fast R-CNN enables end-to-end detector training on shared convolutional features and shows compelling accuracy and speed (Gkioxari et al., 2015).
- Faster R-CNN (Ren et al. 2015)
  - Deep fully convolutional network that proposes regions
  - Fast R-CNN detector that uses the proposed regions.

Objects Detection - You Only Look Once (YOLO)

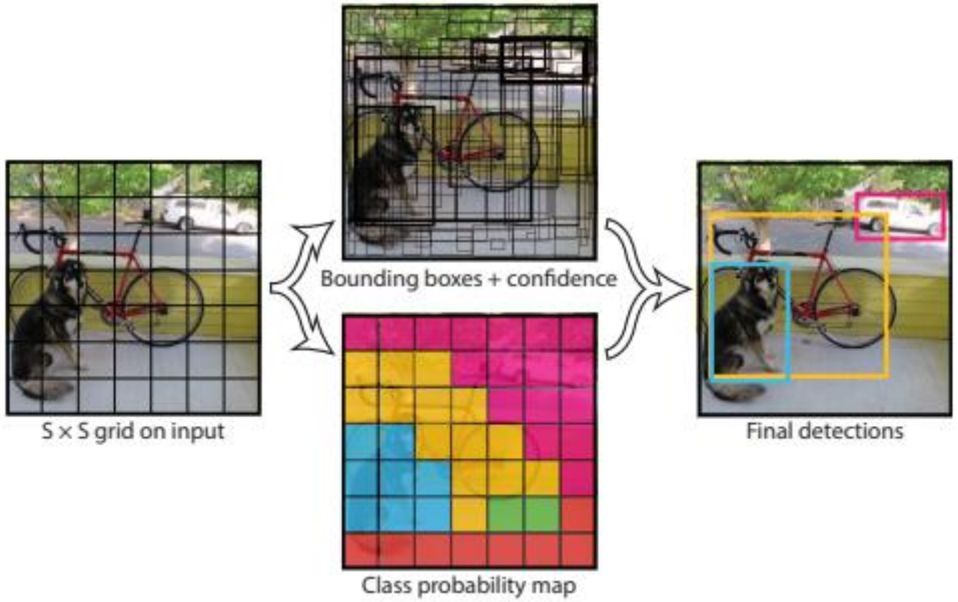


Figure 2: The Model. Our system models detection as a regression problem. It divides the image into an $S \times S$ grid and for each grid cell predicts B bounding boxes, confidence for those boxes, and C class probabilities. These predictions are encoded as an $S \times S \times (B * 5 + C)$ tensor.

Redmon et al.(2016)

- abandons “proposal detection + verification” : Network **divides the image into regions and predicts bounding boxes and probabilities** for each region simultaneously.
- Drawbacks: YOLO suffers from a **drop of the localization accuracy, especially for some small objects**, so subsequent versions have paid more attention to this problem (Redmon et al., 2017; Redmon et al., 2018).

YOLO V3
10 fps

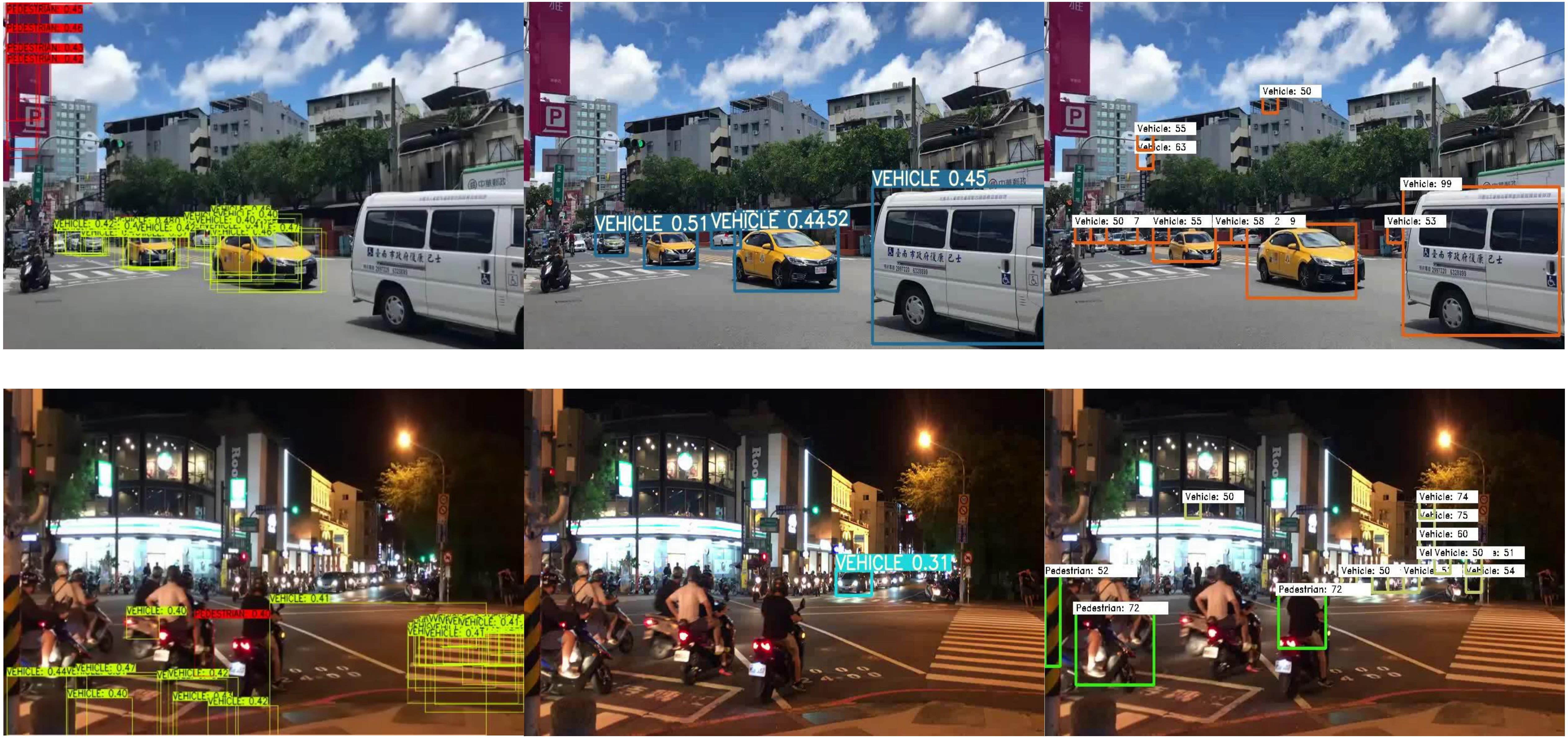
YOLO V3
2.5fps

Experiment

- Frequency of frame fetched & increasing scene diversity; for example :

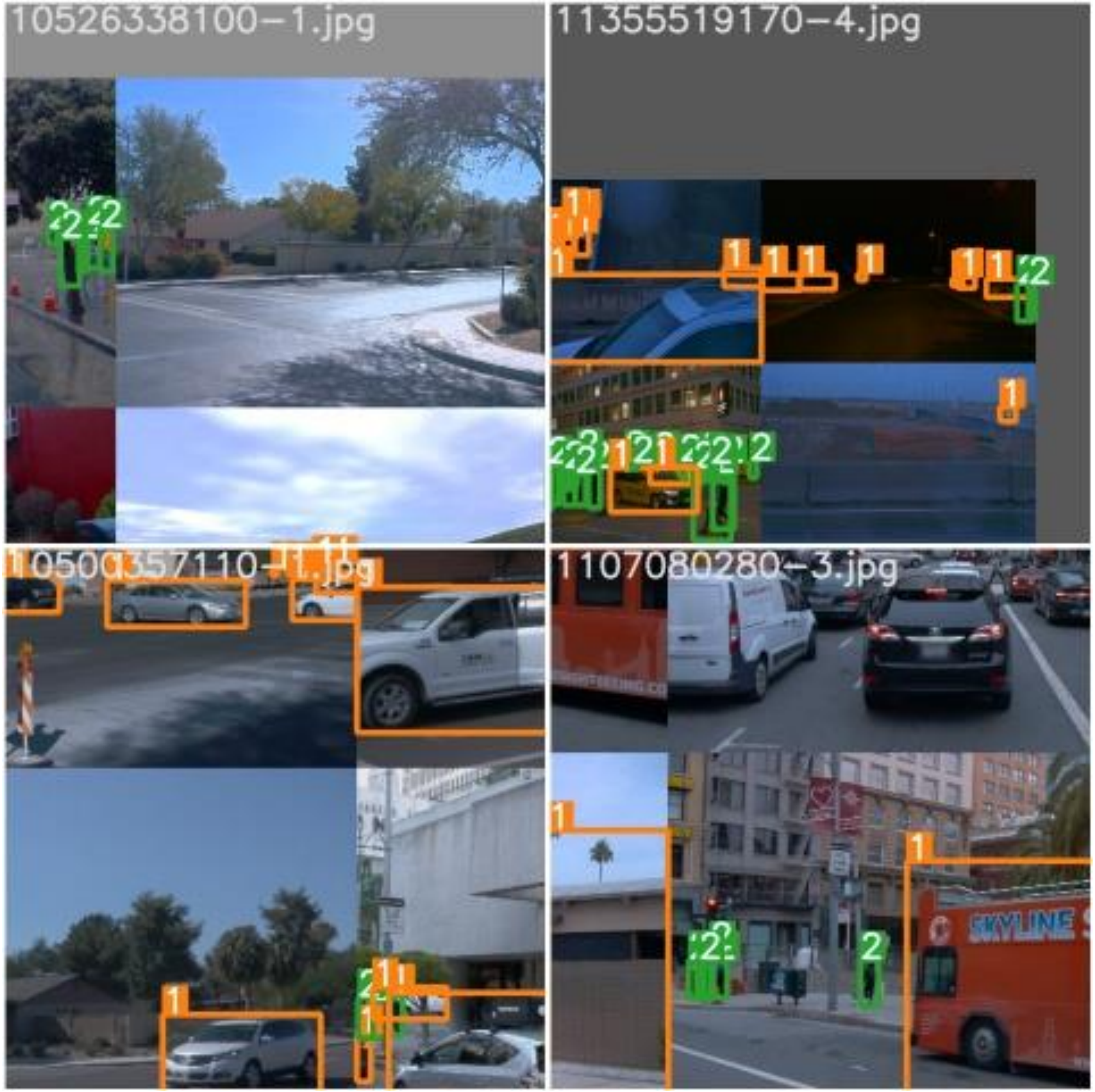
YOLO V3	**10 fps**	**2.5 fps**
Number of image (pieces)	990	3780
mAP (%)	1.16	3.57
AP of Pedestrian(%)	0.17	0.01
AP of Vehicle(%)	3.30	7.12
AP of Cyclist(%)	0.01	

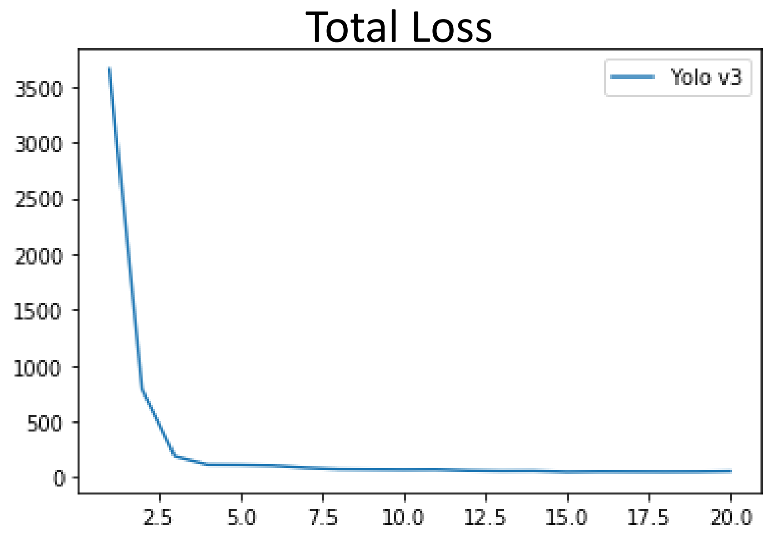
YOLO v3 | YOLO v4 | Faster RCNN

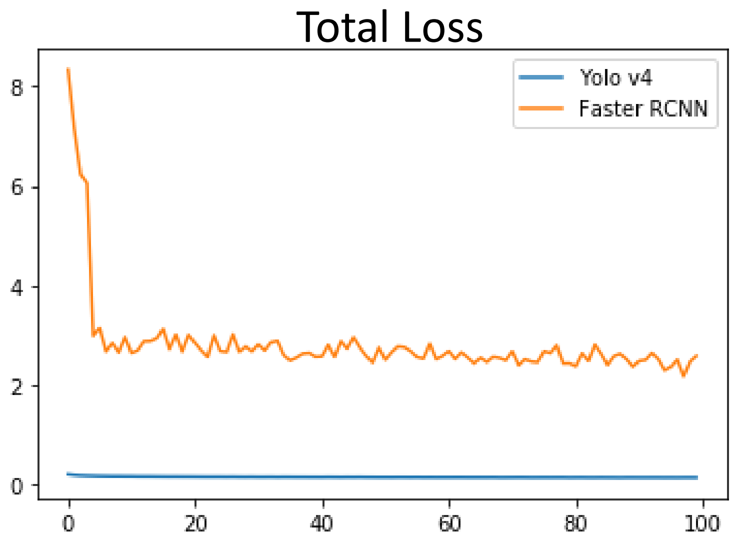


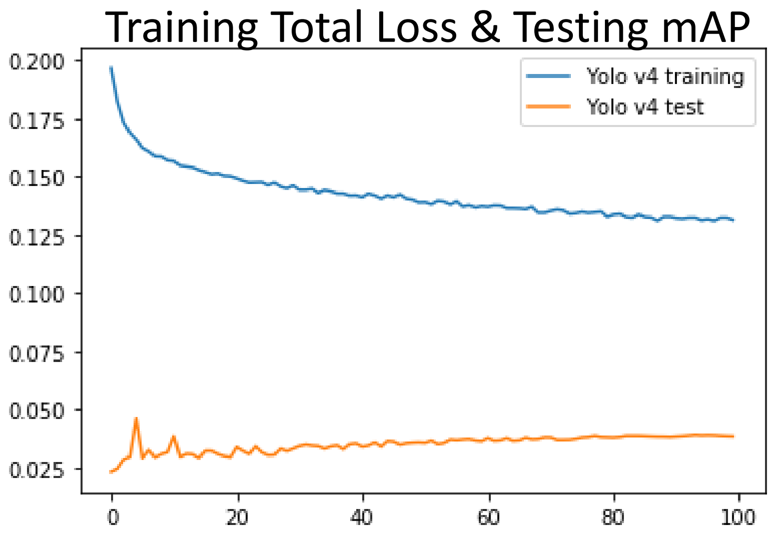
YOLO v4

Different from YOLO v3









Long-tailed Learning Model

Test Dataset	mAP (%)	AP of CYCLIST(%)	AP of PEDESTRIAN(%)	AP of VEHICLE(%)
Faster - RCNN	48.85		55.7	42.0
YOLO v3	3.56		0.01	7.12
YOLO v4	0.04			

Challenge

- Version Problem

Test	Python Version	Learning Framework
Faster - RCNN	3.6	TensorFlow-GPU 2.1.0
YOLO v3	3.6	TensorFlow 1.14
YOLO v4	3.7	PyTorch 1.5

- Isolated Environment

- Imbalanced Dataset
- Overlapped Region Proposals
- Dataset Diversity and Volume
- Metric Understanding

- Collecting More / Data Augmentation
- Tuning Intersection of Union (IoU)
- Data Enforcement

Thanks

Contributors :周禮宏, 唐郁秀,紀昕妤, 賴煜翔和顏振宇 (*sorted by strokes of last name*)