2019/12/19

Vehicle Detection Based on UAV Video and Transfer Learning

By: Runze Yuan

Machine Learning Application

CONTENTS

UAV Detection
The Background

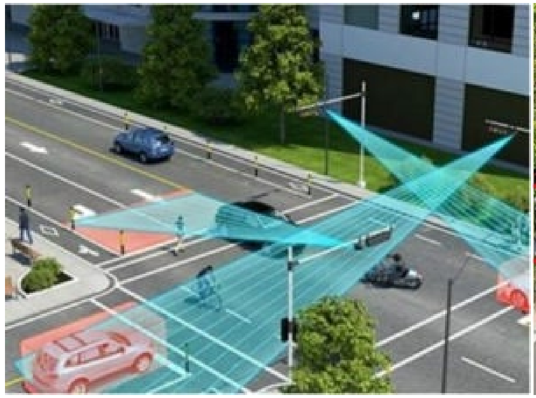
01

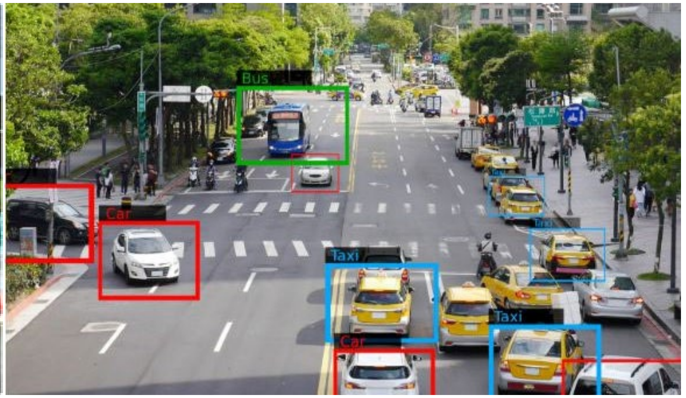
UAV Detection

PART 01

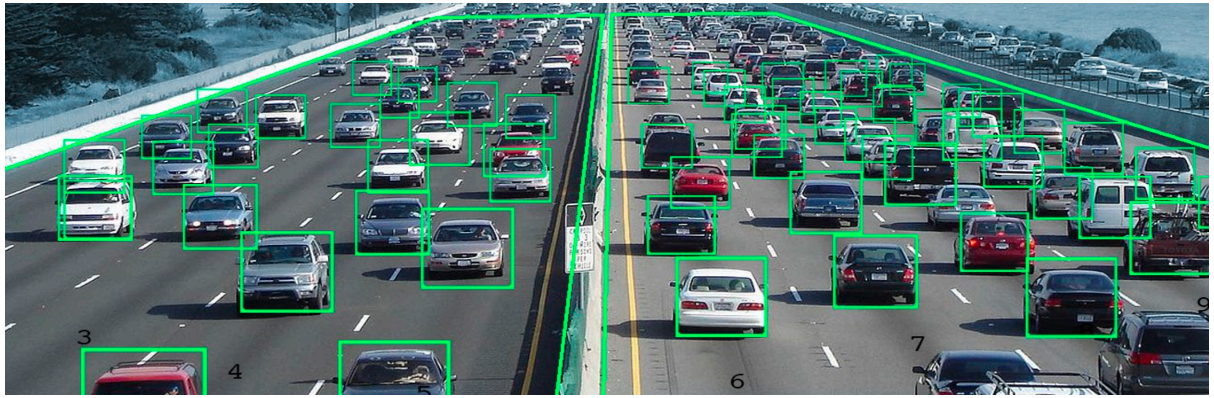
BACKGROUND

Usually, the traffic management department chooses to use a fixed-view CCTV system to monitor the vehicles on the road, to count the number of vehicles, and to find traffic accidents on the road. Sometimes this system is also used to track the target vehicle in criminal incidents. Some corresponding vehicle detection algorithms have also been developed to implement the functions described above. However, the existing fixed-view CCTV system has some problems.





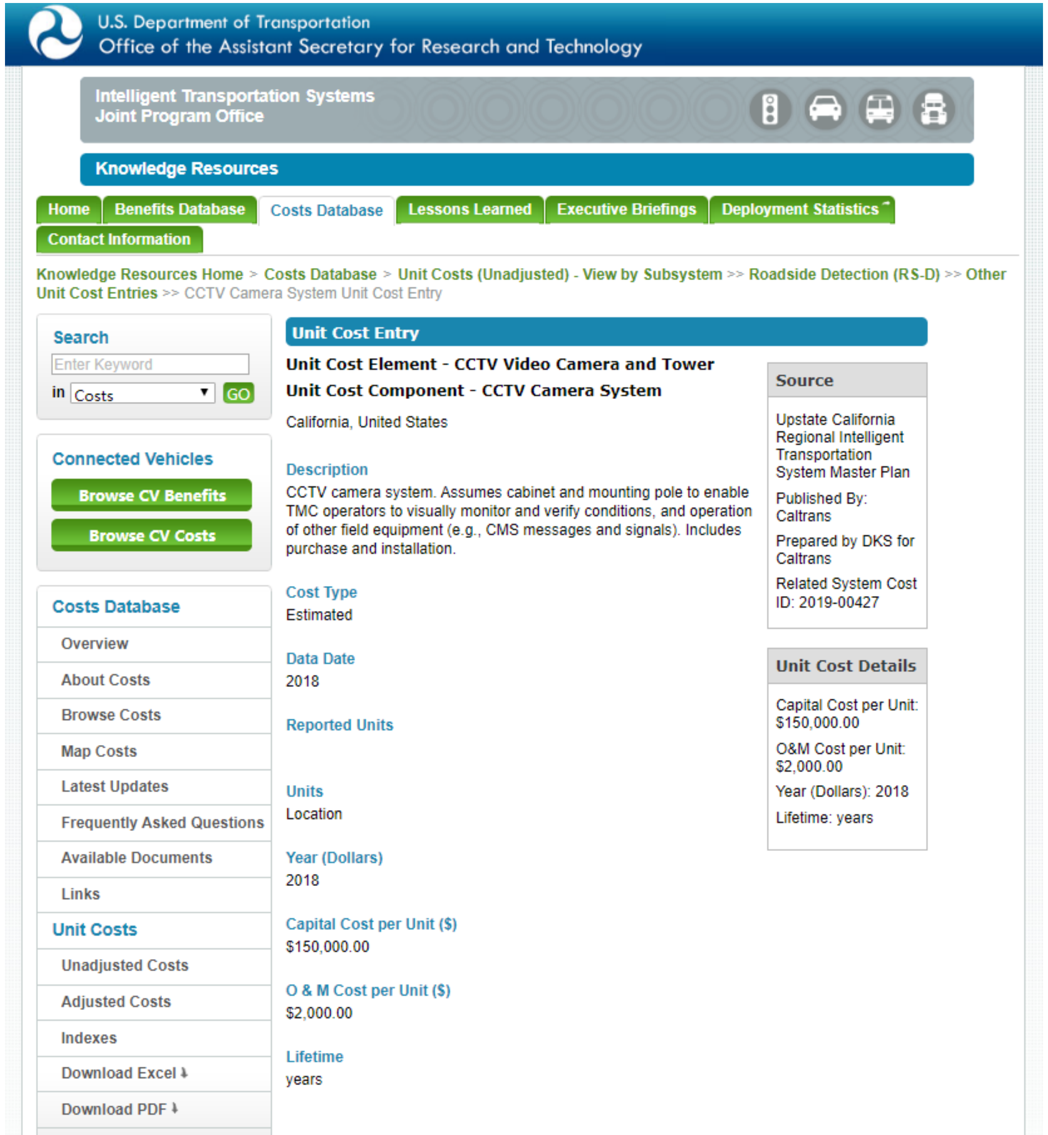
A Fixed-angle Camera



BACKGROUND

Usually, the traffic management department chooses to use a fixed-view CCTV system to monitor the vehicles on the road, to count the traffic flow, and to find traffic accidents on the road. Sometimes this system is also used to track the target vehicle in order to deal with criminal incidents. Some corresponding vehicle detection algorithms have also been developed to implement the functions described above. However, the existing fixed-view CCTV system has some drawbacks.

A Fixed-angle Camera



BACKGROUND

An unmanned aerial vehicle (UAV) (or uncrewed aerial vehicle, commonly known as a drone) is an aircraft without a human pilot on board and a type of unmanned vehicle. UAVs are a component of an unmanned aircraft system (UAS); which include a UAV, a ground-based controller, and a system of communications between the two. The flight of UAVs may operate with various degrees of autonomy: either under remote control by a human operator or autonomously by onboard computers.

Unmanned Aerial Vehicles are increasingly being used in surveillance and traffic monitoring thanks to their high mobility and ability to cover areas at different altitudes and locations.

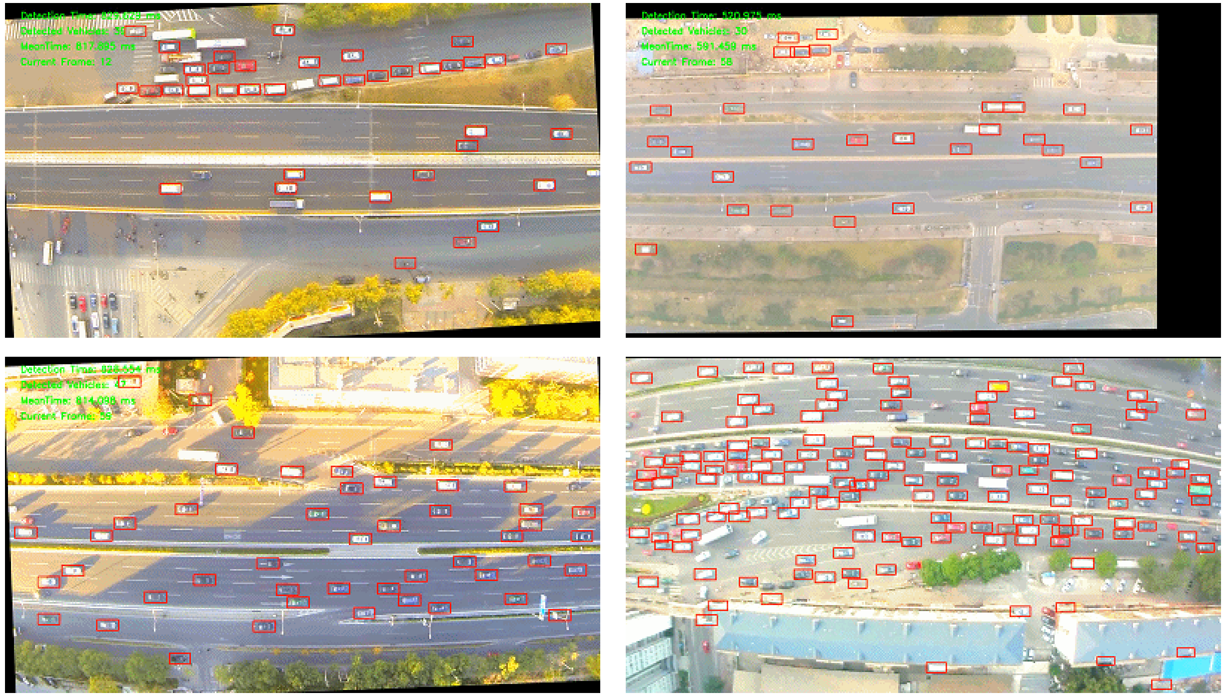
Price: USD $ 1790

https://en.wikipedia.org/wiki/Unmanned_aerial_vehicle

BACKGROUND

An unmanned aerial vehicle (UAV) (or uncrewed aerial vehicle, commonly known as a drone) is an aircraft without a human pilot on board and a type of unmanned vehicle. UAVs are a component of an unmanned aircraft system (UAS); which include a UAV, a ground-based controller, and a system of communications between the two. The flight of UAVs may operate with various degrees of autonomy: either under remote control by a human operator or autonomously by onboard computers.

Unmanned Aerial Vehicles are increasingly being used in surveillance and traffic monitoring thanks to their high mobility and ability to cover areas at different altitudes and locations.



Results of UAV Detection System

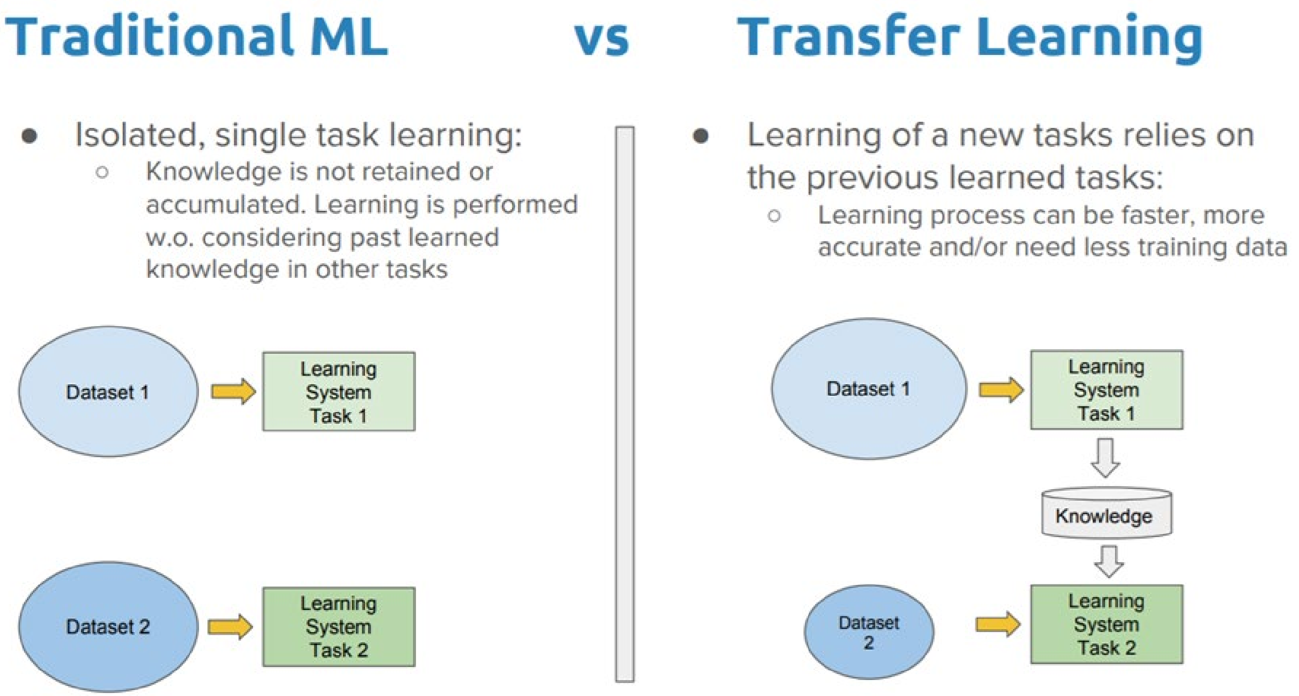
https://en.wikipedia.org/wiki/Unmanned_aerial_vehicle
http://www.xinkaiwu.com/uav.html

Transfer Learning
The Definition

02

Transfer Learning PART 02

What Is Transfer Learning?



Transfer learning is a machine learning method where a model developed for a task is reused as the starting point for a model on a second task.

By using transfer learning, we can train a model of pretty good performance without training from the very beginning, which would take a lot of time and need thousands of labeled pictures.

In this way, we can train a model to detect vehicles in the UAV videos with a relatively smaller dataset.

https://towardsdatascience.com/a-comprehensive-hands-on-guide-to-transfer-learning-with-real-world-applications-in-deep-learning-212bf3b2f27a

YOLOv3 Model

Short Introduction

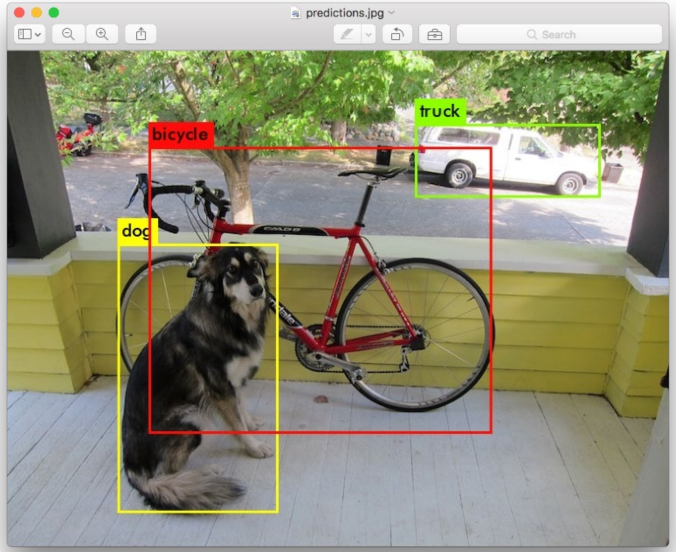
03

YOLOv3 Model PART 03

YOLOv3:Real-Time Object Detection

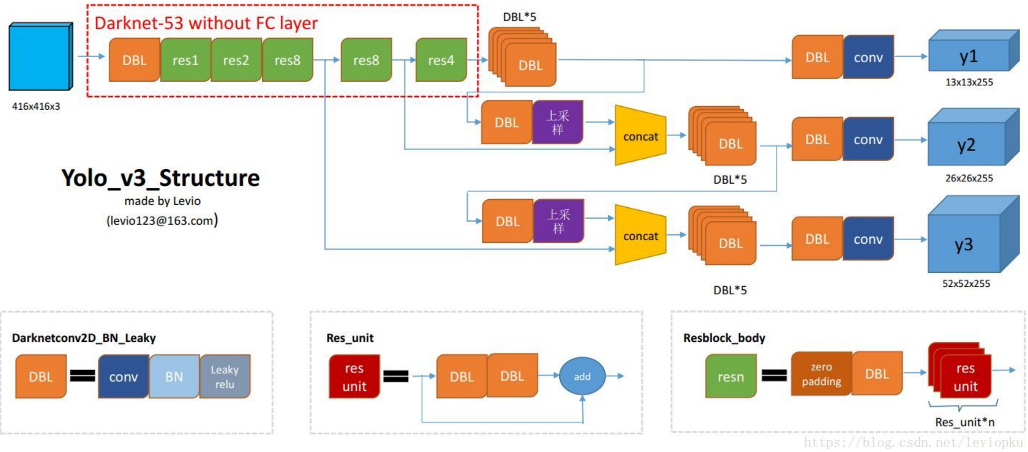
Darknet-53: the core of YOLOv3

	Type	Filters	Size	Output
	Convolutional	32	3 3	256 256
	Convolutional	64	3 3 / 2	128 128
1	Convolutional	32	1 1	
	Convolutional	64	3 3	
	Residual			128 128
	Convolutional	128	3 3 / 2	64 64
2	Convolutional	64	1 1	
	Convolutional	128	3 3	
	Residual			64 64
	Convolutional	256	3 3 / 2	32 32
8	Convolutional	128	1 1	
	Convolutional	256	3 3	
	Residual			32 32
	Convolutional	512	3 3 / 2	16 16
8	Convolutional	256	1 1	
	Convolutional	512	3 3	
	Residual			16 16
	Convolutional	1024	3 3 / 2	8 8
4	Convolutional	512	1 1	
	Convolutional	1024	3 3	
	Residual			8 8
	Avgpool		Global	
	Connected		1000	
	Softmax			



You only look once (YOLO) is a state-of-the-art, real-time object detection system.

Based on the structure of YOLOv3, I used the method of transfer learning to train the model using the dataset of VisDrone 2018.



https://pjreddie.com/darknet/yolo/

YOLOv3:Real-Time Object Detection

Microsoft COCO dataset

What is COCO?

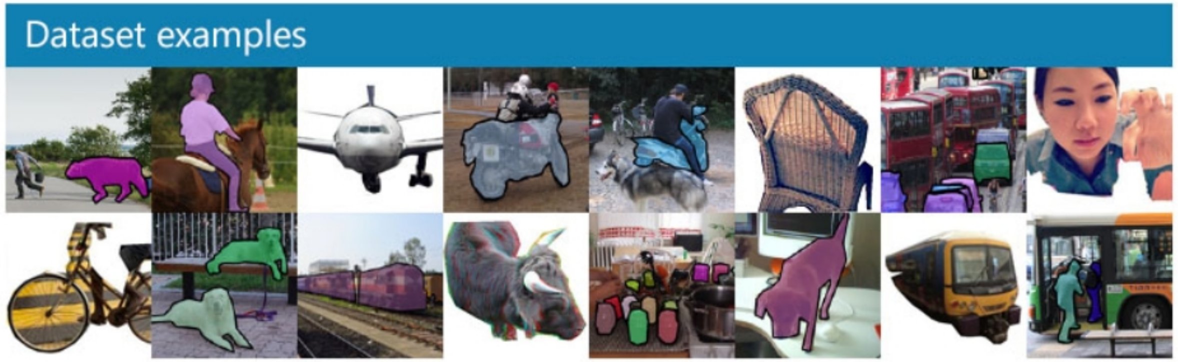
COCO is a large-scale object detection, segmentation, and captioning dataset. COCO has several features:

- Object segmentation
- Recognition in context
- Superpixel stuff segmentation
- 330K images (>200K labeled)
- 1.5 million object instances
- 80 object categories
- 91 stuff categories
- 5 captions per image
- 250,000 people with keypoints

http://cocodataset.org/#home

You only look once (YOLO) is a state-of-the-art, real-time object detection system.

According to the tech report of the author, the original network of YOLO is trained with Microsoft COCO dataset.



Training Dataset

VisDrone2018

04

Training Dataset

PART 04

Vision Meets Drones

VisDrone 2018 Dataset

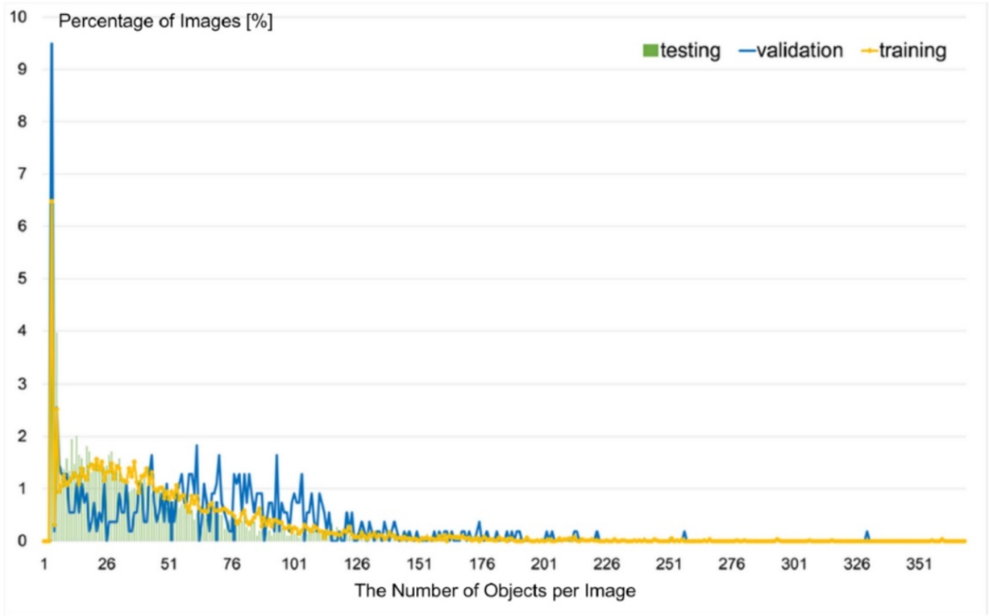
The VisDrone2018 dataset is collected by the AISKYEYE team at Lab of Machine Learning and Data Mining , Tianjin University, China.

The benchmark dataset consists of 288 video clips formed by 261,908 frames and 10,209 static images, captured by various drone-mounted cameras, covering a wide range of aspects including location (taken from 14 different cities separated by thousands of kilometers in China), environment (urban and country), objects (pedestrian, vehicles, bicycles, etc.), and density (sparse and crowded scenes).

http://www.aiskyeye.com/

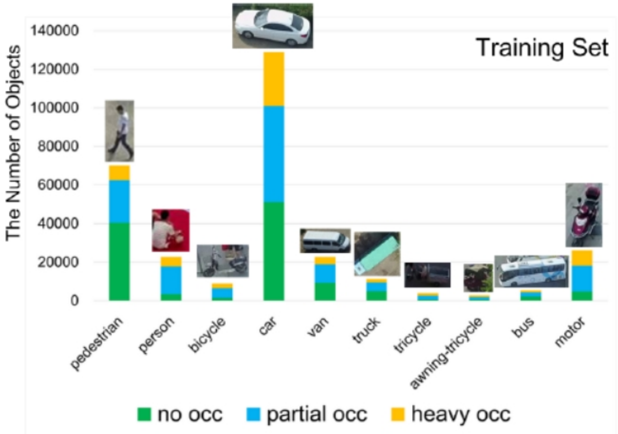
Vision Meets Drones

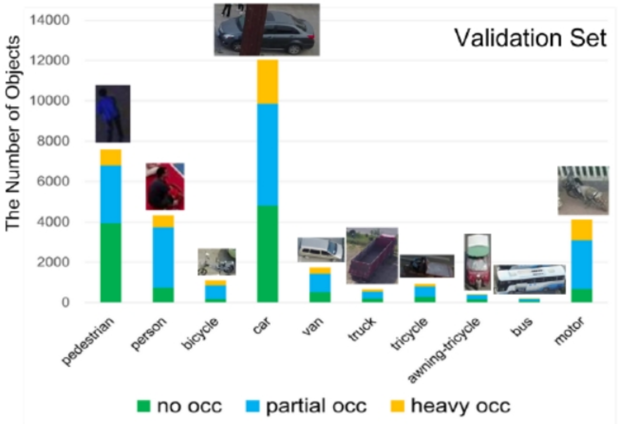
VisDrone 2018 Dataset



The model is trained with VisDrone-2018-Dataset-Train (1.44GB), using VisDrone-2018-Dataset-Val (0.07GB) as validation set.

VisDrone-2018-Dataset-Train contains 6287 pictures;
VisDrone-2018-Dataset-Val contains 543 pictures.





http://www.aiskyeye.com/

Training & Results

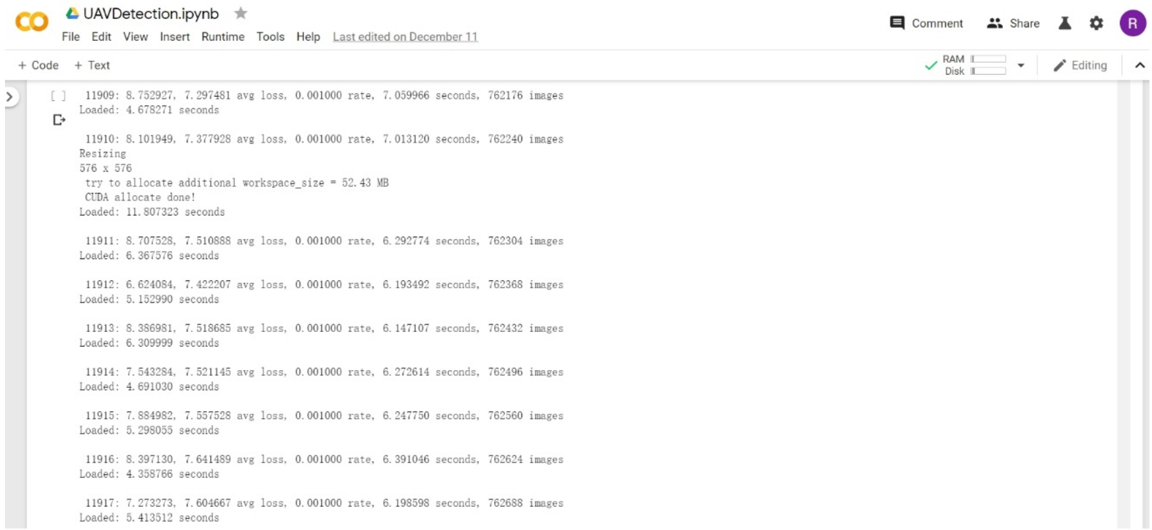
The Current Results

05

Training & Results

PART 05

Training Process



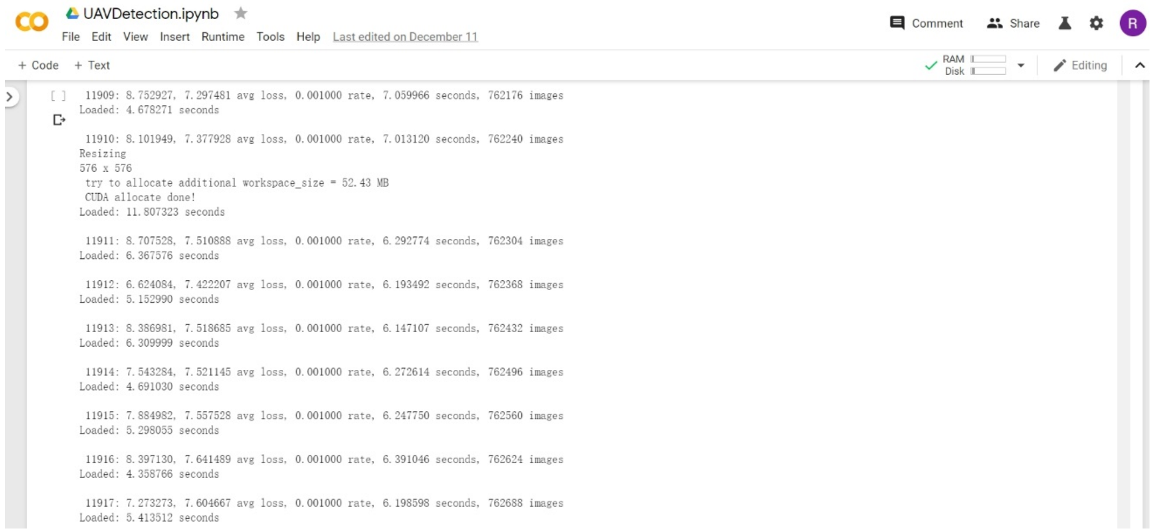
· Trained with VisDrone-2018-Dataset-Train using VisDrone-2018-Dataset-Val as validation set.

· Using Google Colab as the training platform to run the codes. During the training process, the weight file will be saved after 1000 epochs automatically.

· Using the Google Drive as the VM drive to save the training data and trained weights of model.

· The files in my Google Drive will be synced to my laptop.

Training Process

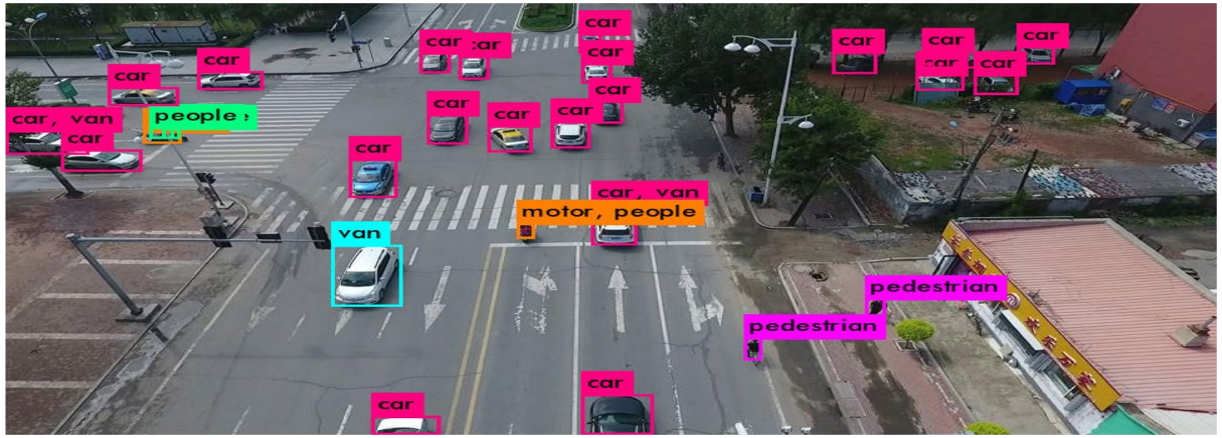


· Trained with VisDrone-2018-Dataset-Train using VisDrone-2018-Dataset-Val as validation set.

· Trained the first model using 10 kinds of objects for about 72 hours with 25000 epochs.
The avg. loss is about 12.

· Trained the second model using 5 kinds of vehicle: car, bus, van, truck, motor. About 30 hours, with 18000 epochs. The avg. loss is about 7.7.

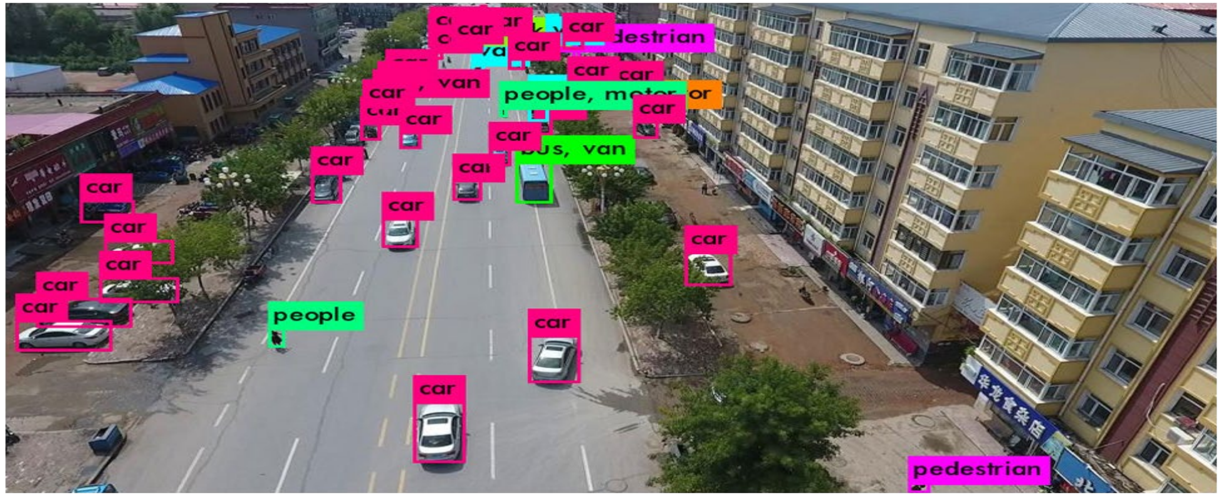
Experiment Results

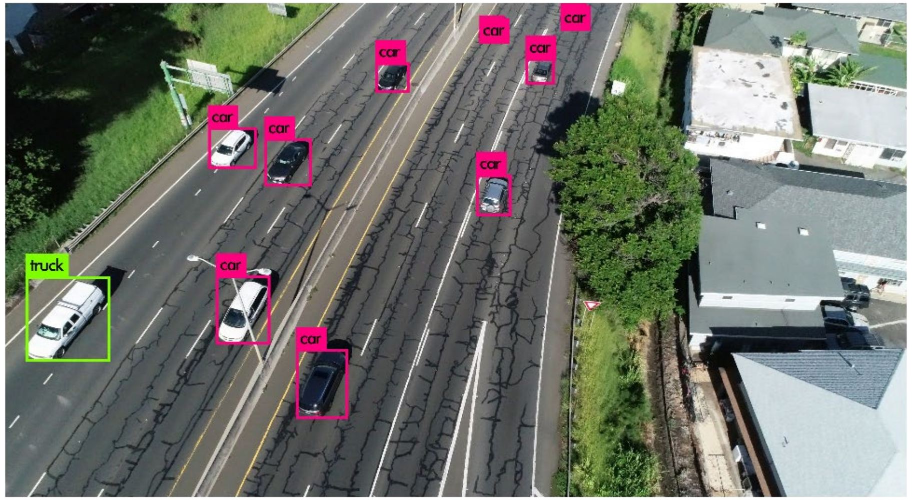


- Trained with VisDrone-2018-Dataset-Train using VisDrone-2018-Dataset-Val as validation set with 10 kinds of objectives.

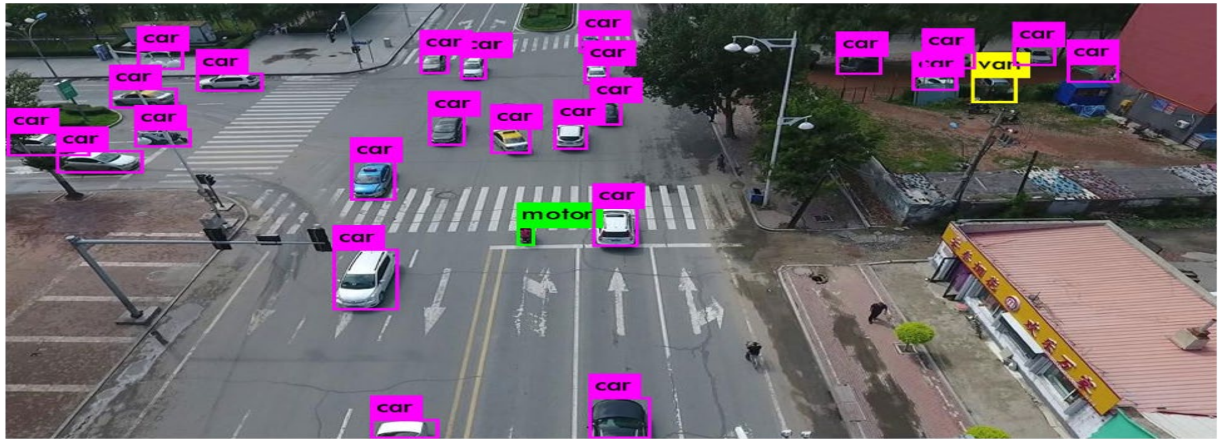
- About 72 hours with 25000 epochs.

- The avg. loss is about 12.





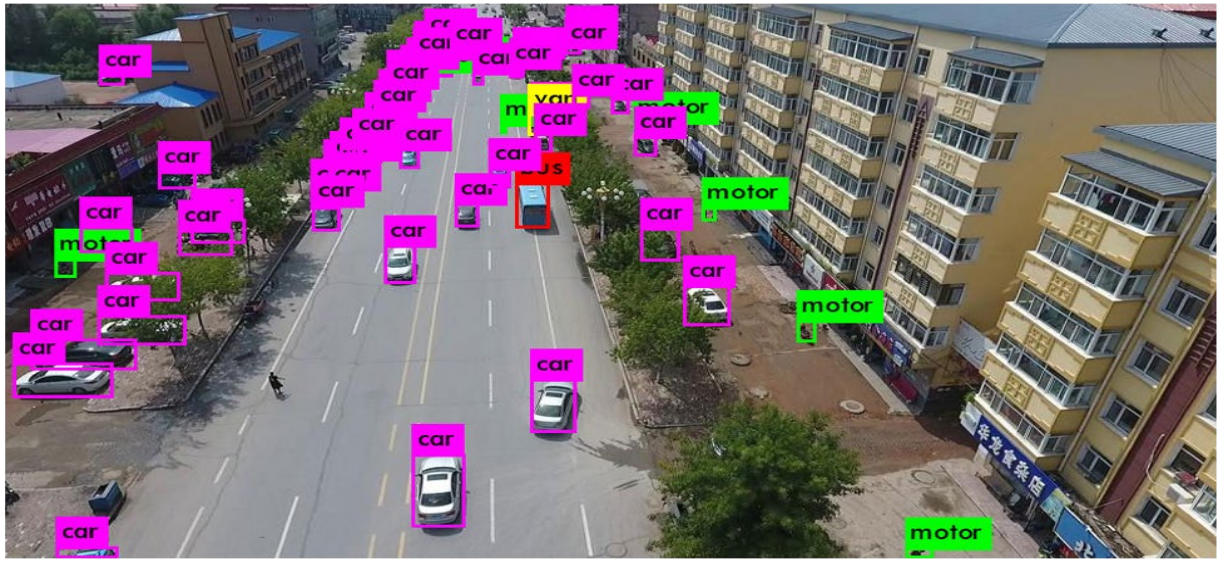
Experiment Results

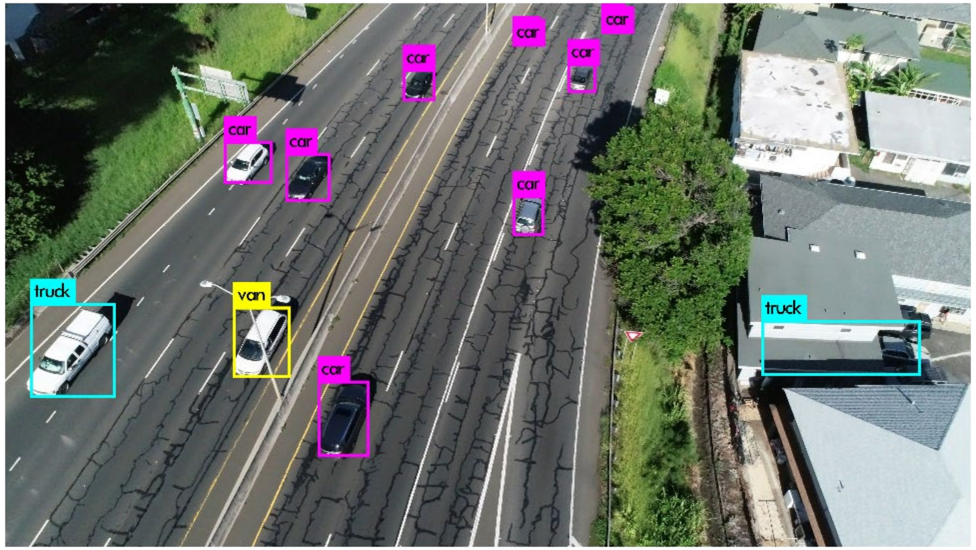


· Trained with VisDrone-2018-Dataset-Train using VisDrone-2018-Dataset-Val as validation set but using vehicles only.

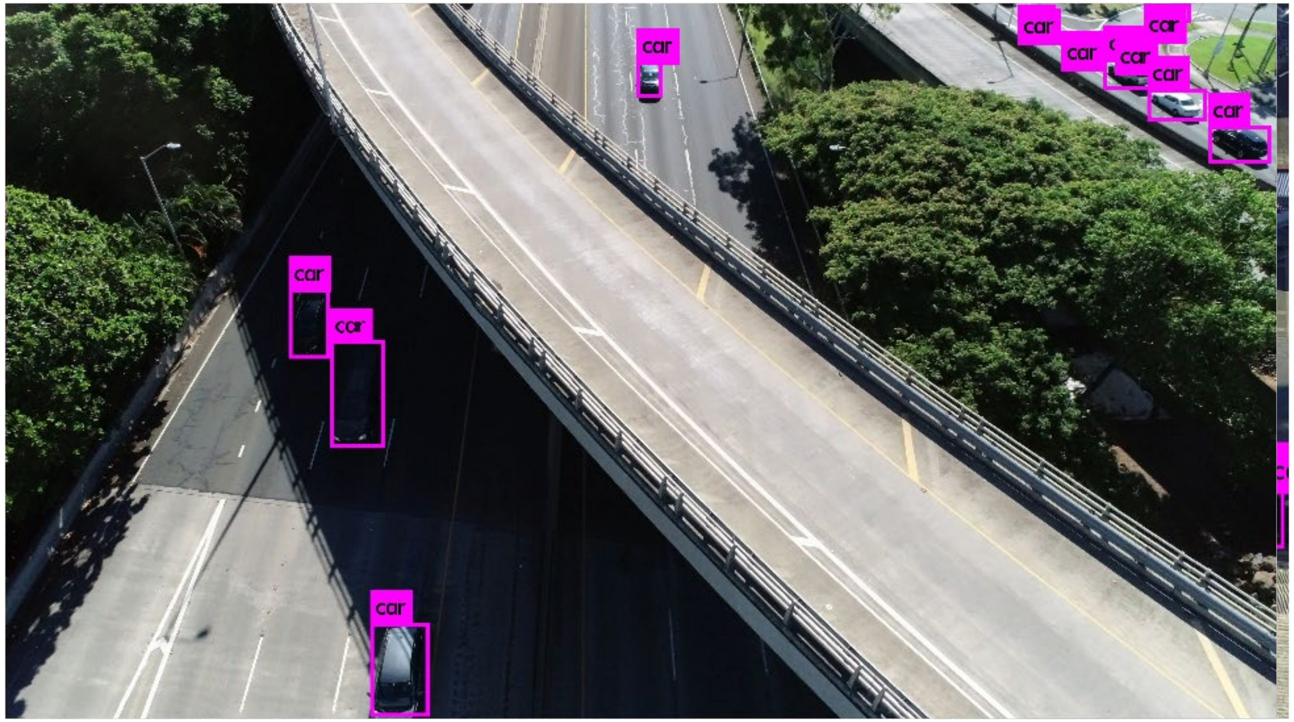
· About 30 hours with 18000 epochs.

· The avg. loss is about 7.7.





Experiment Results



· Trained with VisDrone-2018-Dataset-Train using VisDrone-2018-Dataset-Val as validation set but using vehicles only.

· About 30 hours with 18000 epochs.

· The avg. loss is about 7.7.

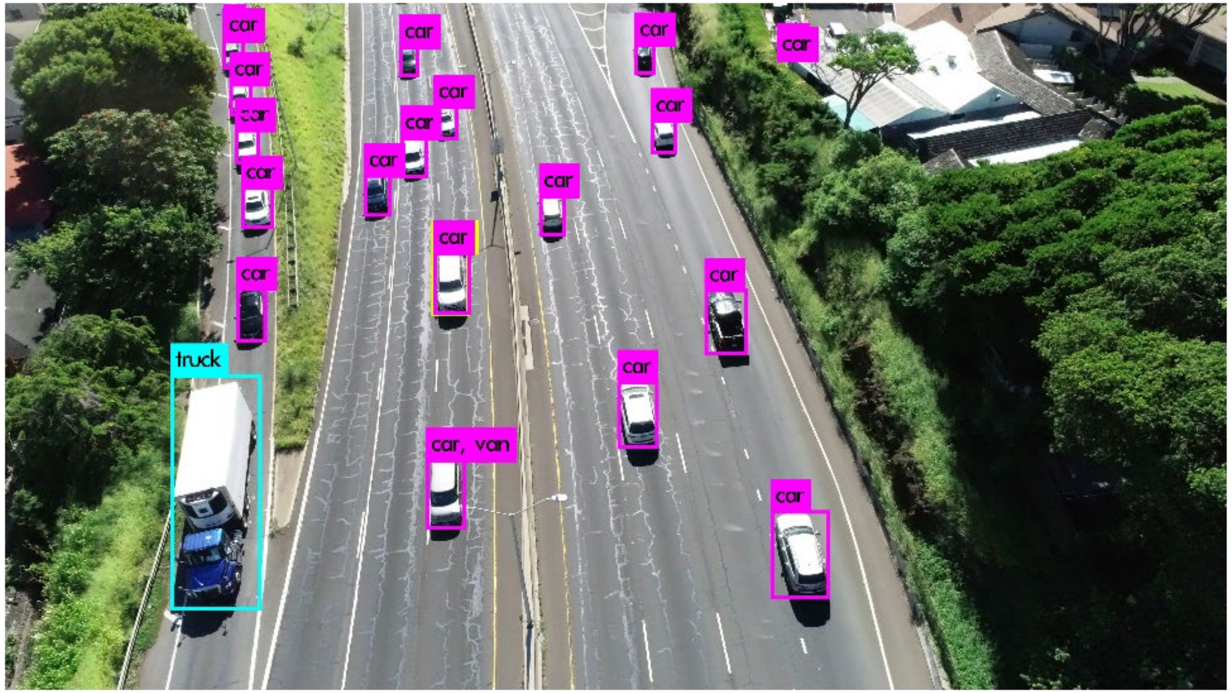
Future Work

To Build A Robust System

06

Future Work PART 05

Future Work



The current model is trained with about 7000 pictures in about 25000 epochs but it cannot tell us if one detected vehicle is on road and may give wrong results.

In the next steps, I will train the model with more labeled images of cars, trucks, motorcycles, etc. to make it better.

Based on the detection module, we still need to add a module to calculate the number of vehicles detected and the average speed of the traffic system.

Finally, I want to build a system to detect the vehicles in real-time and help to manage the traffic system automatically.

2019/12/19

THANK YOU!

2019/12/19

Questions?

The End