HAL-Retina-Net: Hybrid attention based low-resolution retina-net

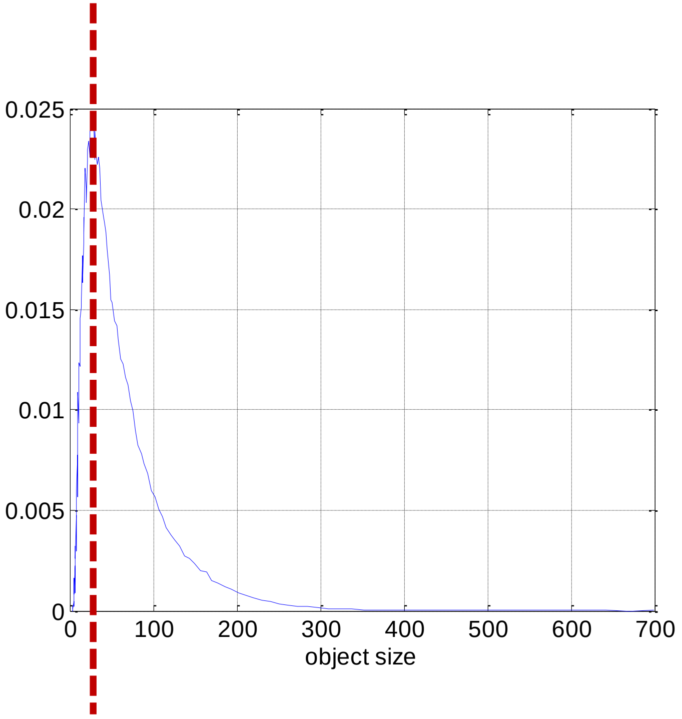
Yali Li, Jianqiang Wang, Zhaoyue Xia, Shengjin Wang
Department of Electronic Engineering
Tsinghua University

Dataset Description

http://www.aiskyeye.com/

Dataset Description

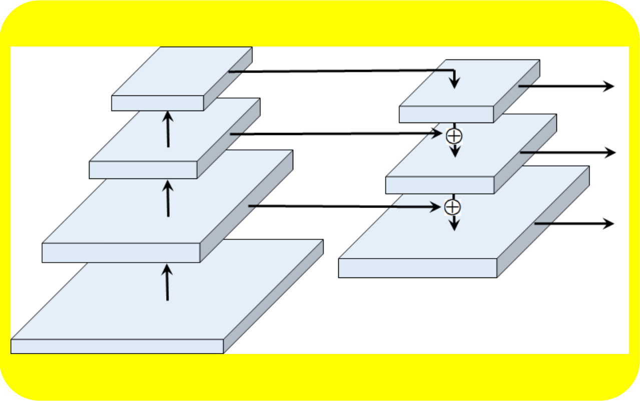
Drones, or general UAVs, equipped with cameras have been fast deployed to a wide range of applications, including agricultural, aerial photography, fast delivery, and surveillance. Consequently, automatic understanding of visual data collected from these platforms become highly demanding, which brings computer vision to drones more and more closely. We are excited to present a large-scale benchmark with carefully annotated ground-truth for various important computer vision tasks, named VisDrone, to make vision meet drones. 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 *263* video clips formed by *179,264* 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).** Note that, the dataset was collected using various drone platforms (*i.e.*, drones with different models), in different scenarios, and under various weather and lighting conditions. These frames are manually annotated with more than *2.5* million bounding boxes of targets of frequent interests, such as ***pedestrians, cars, bicycles, and tricycles***. Some important attributes including scene visibility, object class and occlusion, are also provided for better data utilization.



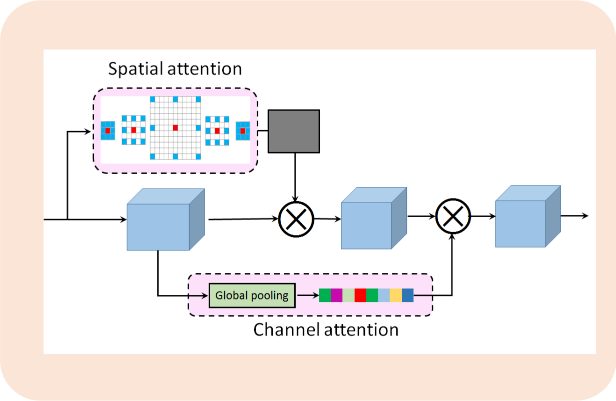
32 pixels

Network Architecture

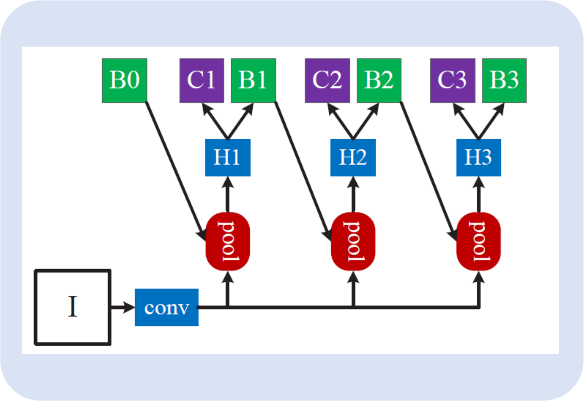
- *We choose Retina-Net as the baseline for this challenge.*
- *To detect low-resolution objects, we remove P6 and P7 from the feature pyramid structure and use the first three pathways, named as P3, P3, P5, for detection.*



Feature pyramid network



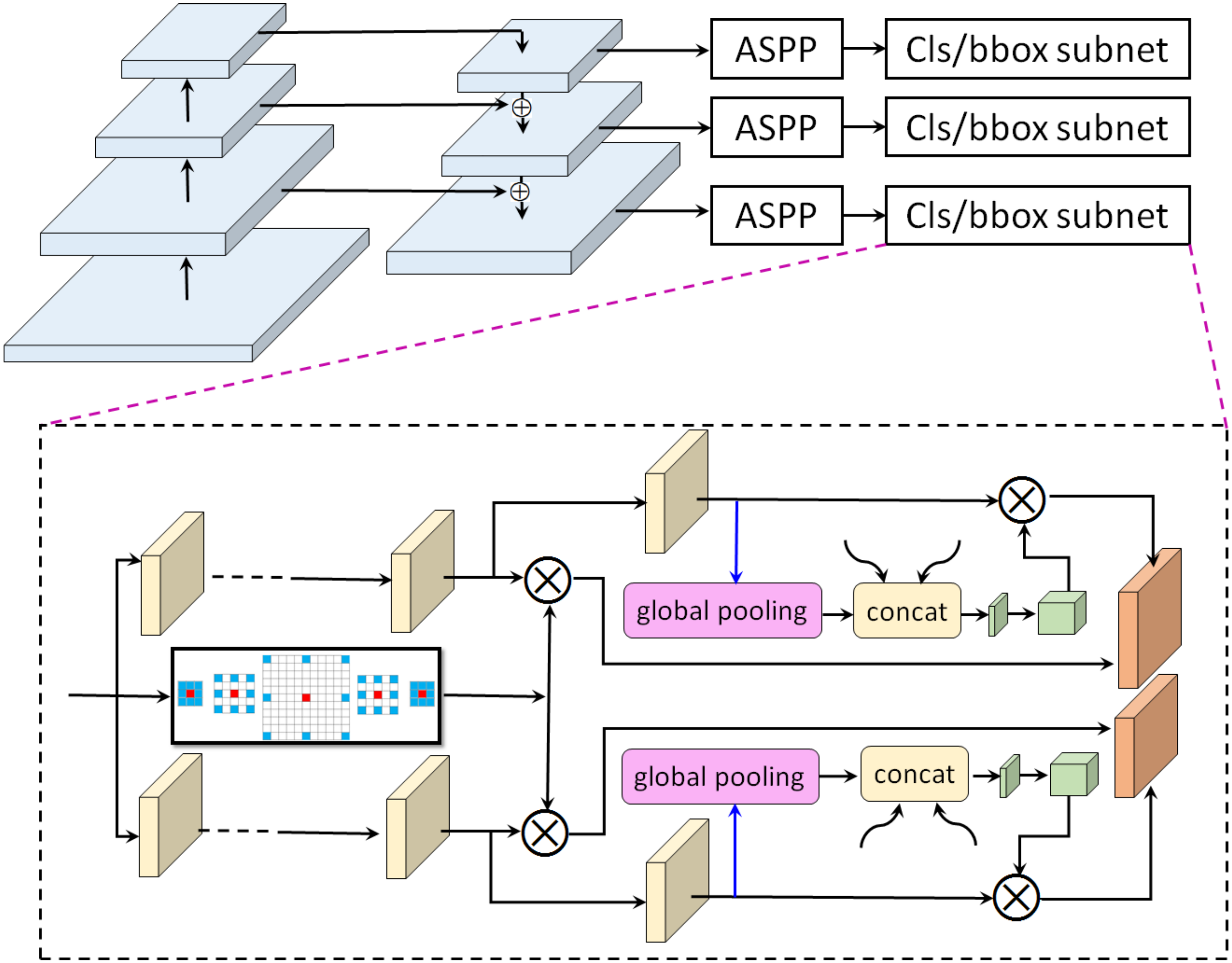
Hybrid attention module



Refinement module with cascade

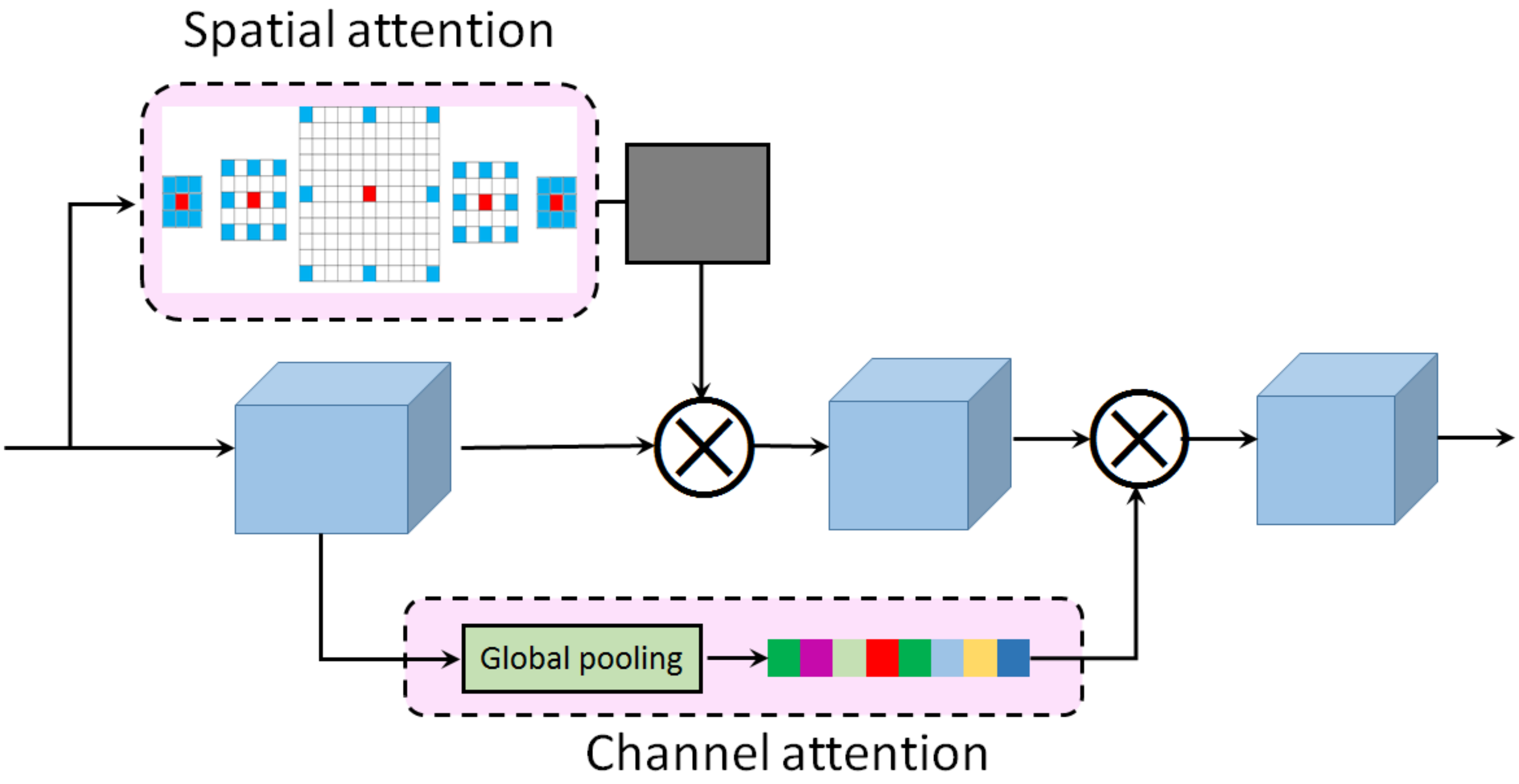
Zhaowei Cai, Nuno Vasconcelos. *Cascade R-CNN: Delving into High Quality Object Detection.CVPR 2018.*

Network Architecture



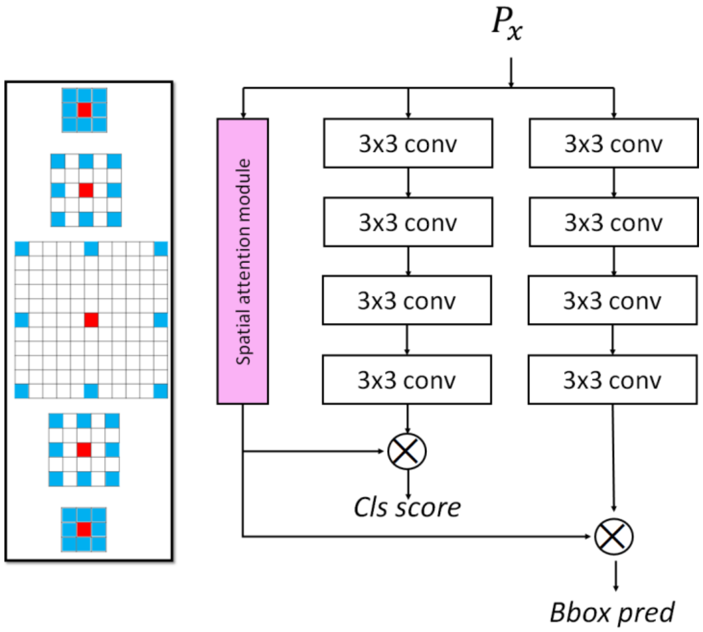
Hybrid Attention

- *Network attention: Spatial attention + channel attention*
- *Learning attention: Multi-scale training*



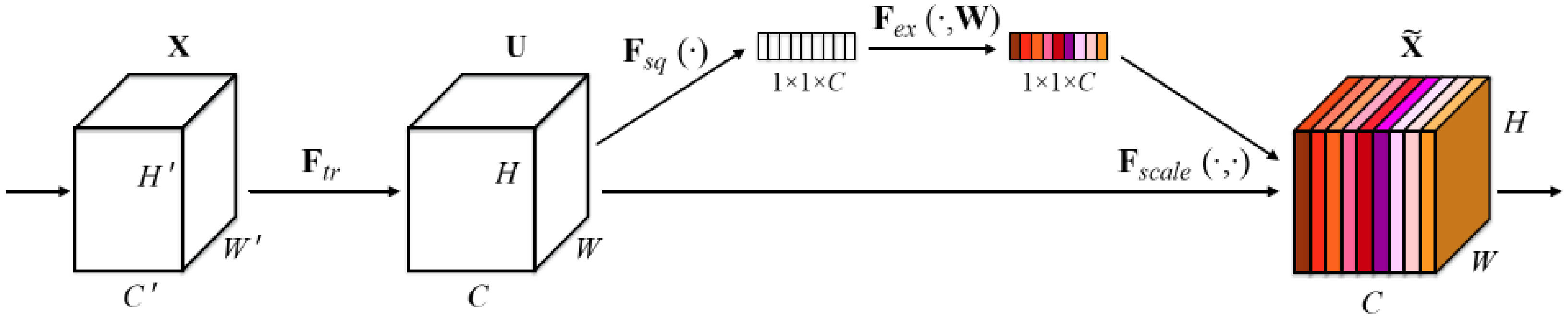
Spatial Attention

- *Use stacked dilated convolution layers to learn the heat maps of interested objects.*
- *Dilation rate needs to be progressively decreased to decrease the gridding mosaic artifacts.*



Channel Attention

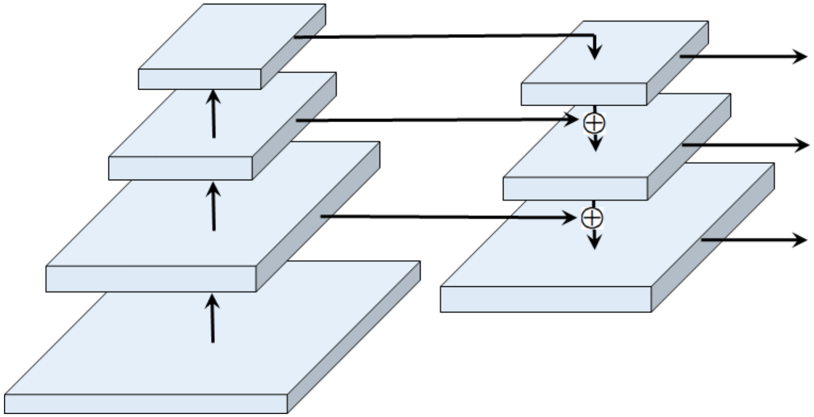
- *Squeeze-and-excitation (SE) modules are used to extract the channel relationship and learn the weights of different channels with the global information.*
- *The global information from different stream to learn the shared channel attention.*

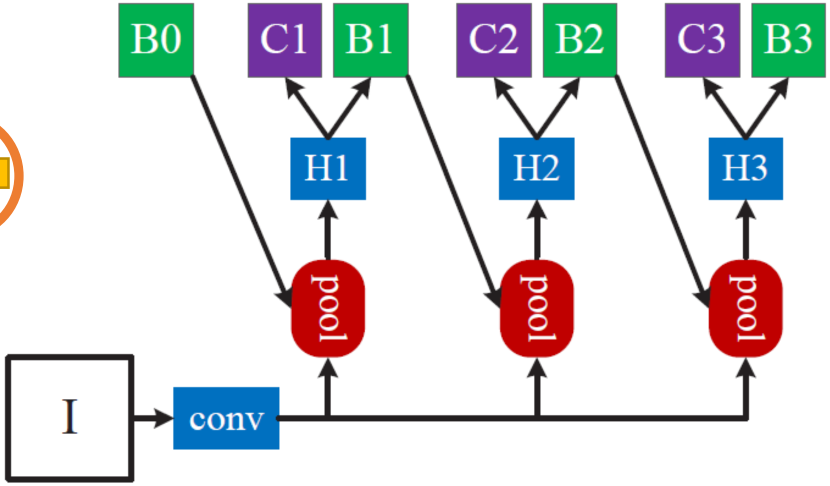


Jie Hu, Li Shen, Gang Sun. Squeeze-and-Excitation Networks. arXiv:1709.01507.

Cascade CNN

- *Iterative box to improve the localization accuracy*
- *Integral loss to improve the detection accuracy.*
- *We add the cascade module for further refinement.*





Zhaowei Cai, Nuno Vasconcelos. Cascade R-CNN: Delving into High Quality Object Detection.CVPR 2018.

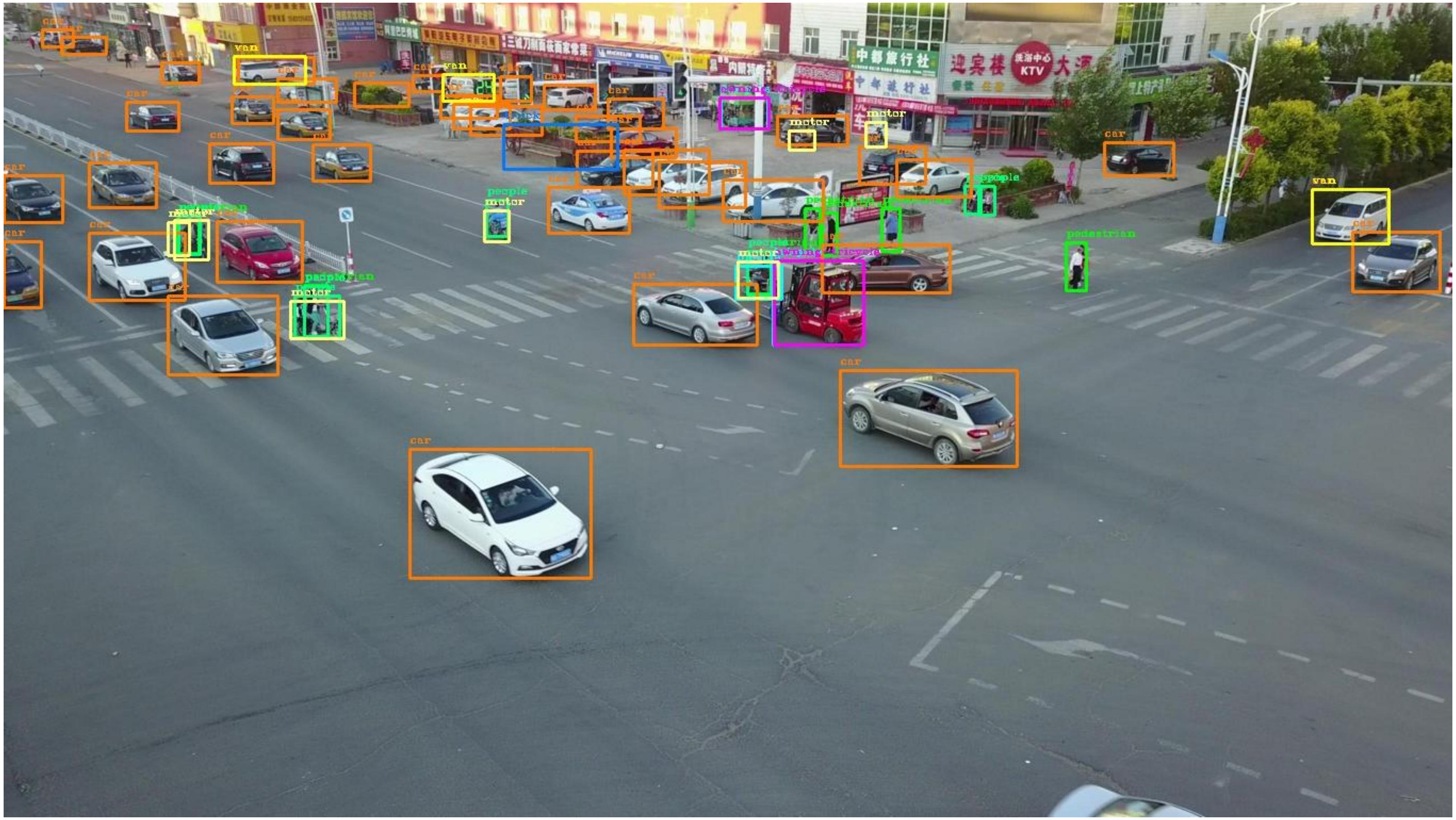
Implementation details

- *Backbone network: SE-ResNeXt-50/SE-ResNeXt-101*
- *Training: firstly trained on COCO[1], with 45.7% mAP, then finetuned with VisDrone2018 train/val set.*
- *Single scale training: learning rate 0.005 of 24k iterations then decreased to 0.0005 with 27k iterations.*
- *Multi-scale training: split the original images into patches (size: 512, 640, 768, 896, 1024), with DOTA devkit [2], nearly 180k sub-images for training.*
- *Weight decay of 0.0001 and momentum of 0.9 are used. The training loss is the sum of focal loss and standard smooth L1 loss used for box regression.*

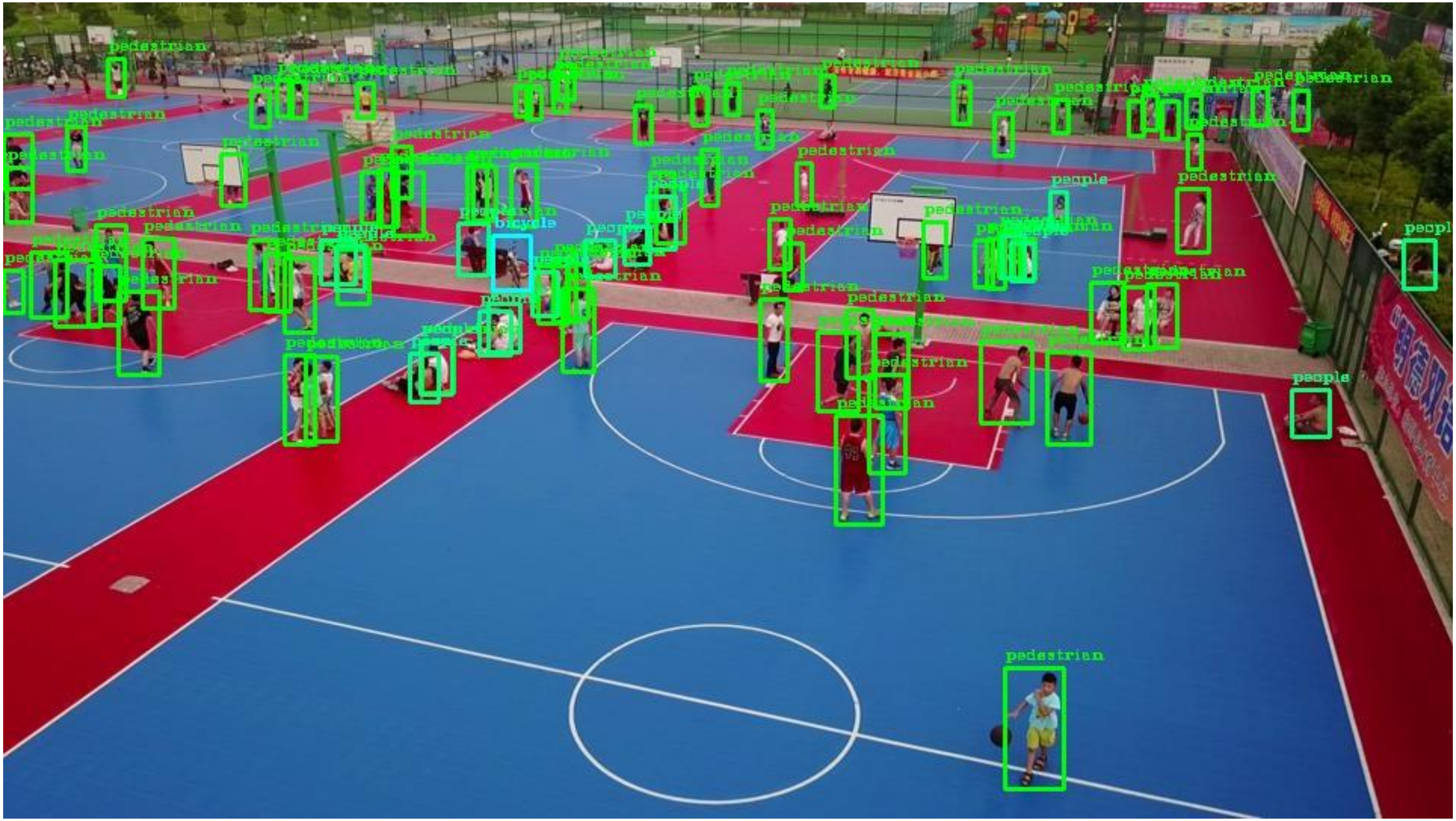
[1] http://cocodataset.org/

[2] http://captain.whu.edu.cn/dotaweb/

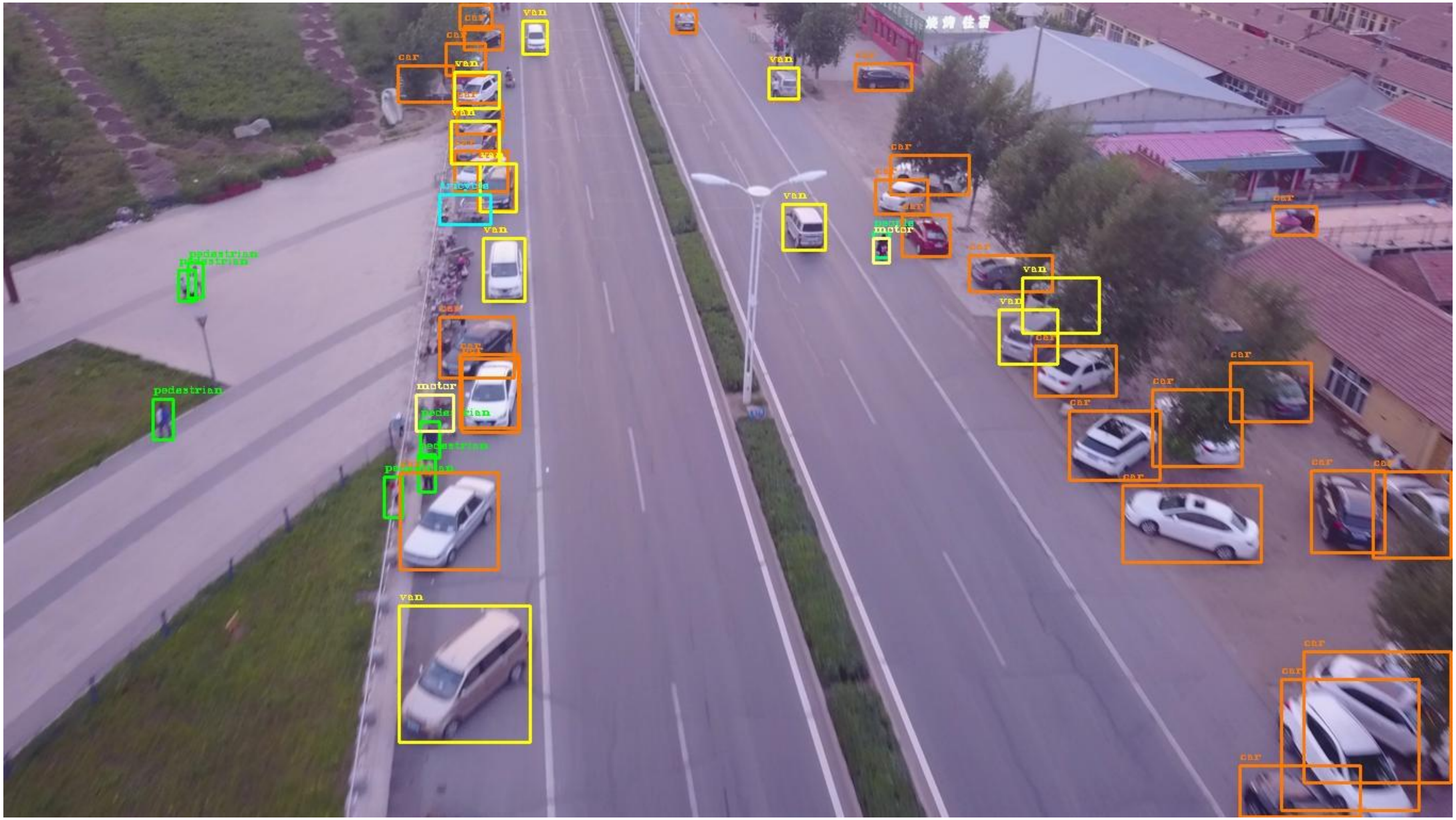
Demos


car
van
truck
motor
people
pedestrian
tricycle
中部旅行社
迎宾楼
KTV

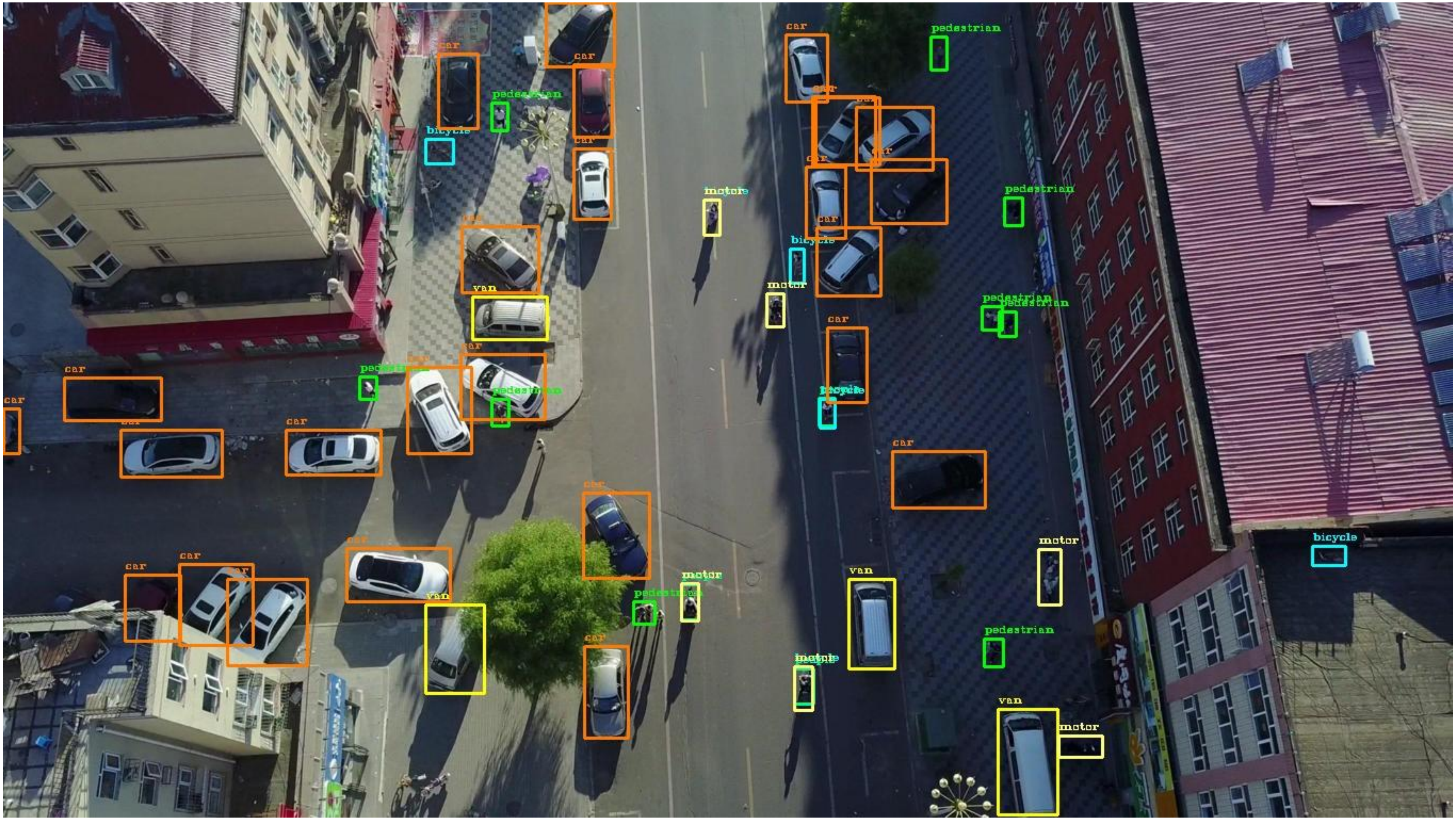
Demos



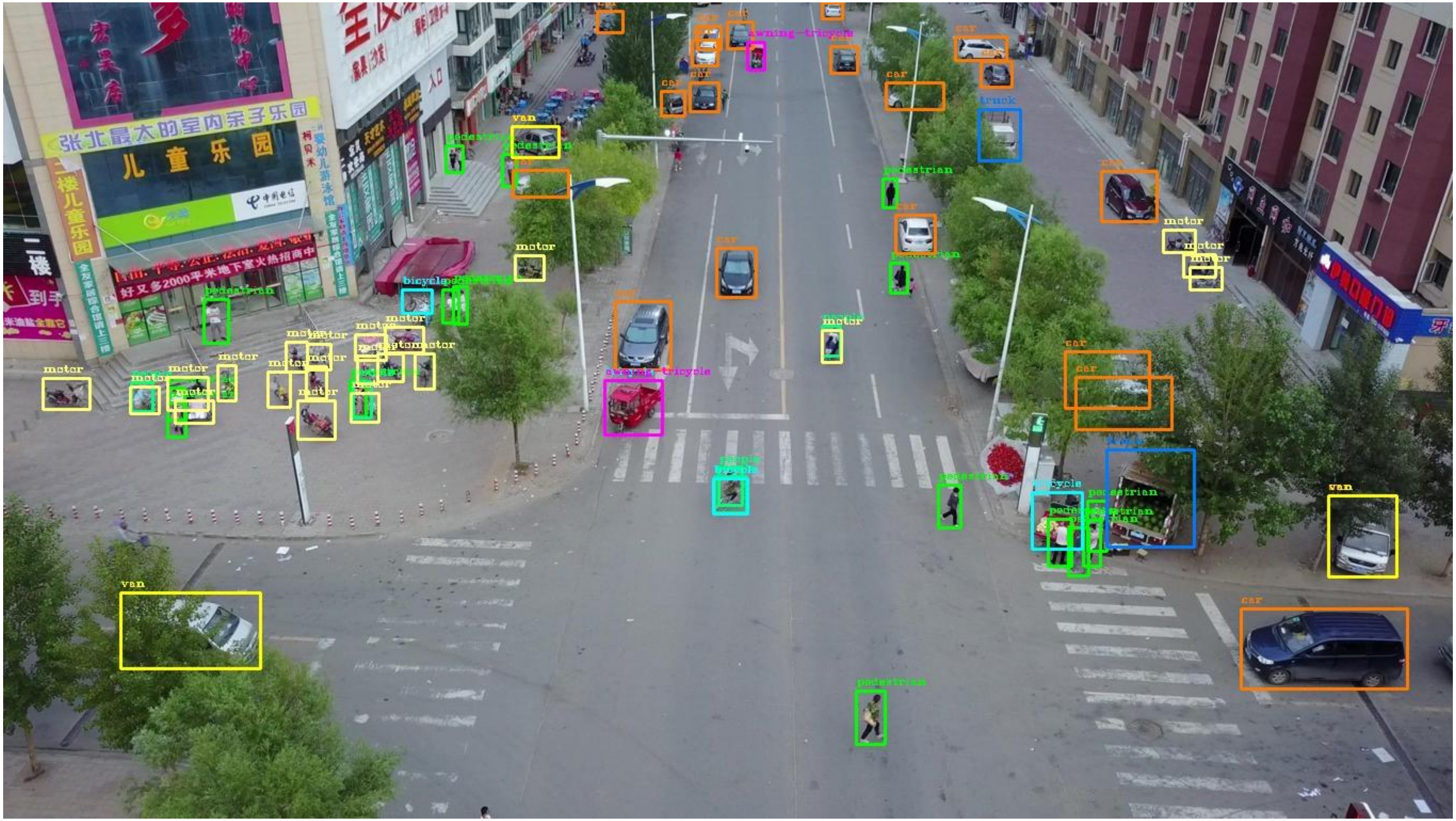
Demos


car
van
car
van
car
van
car
van
van
tricycle
van
van
car
pedestrian
pedestrian
car
car
van
motor
car
van
car
pedestrian
motor
pedestrian
pedestrian
car
car
car
car
car
car
car
car
van
car
car
car
car

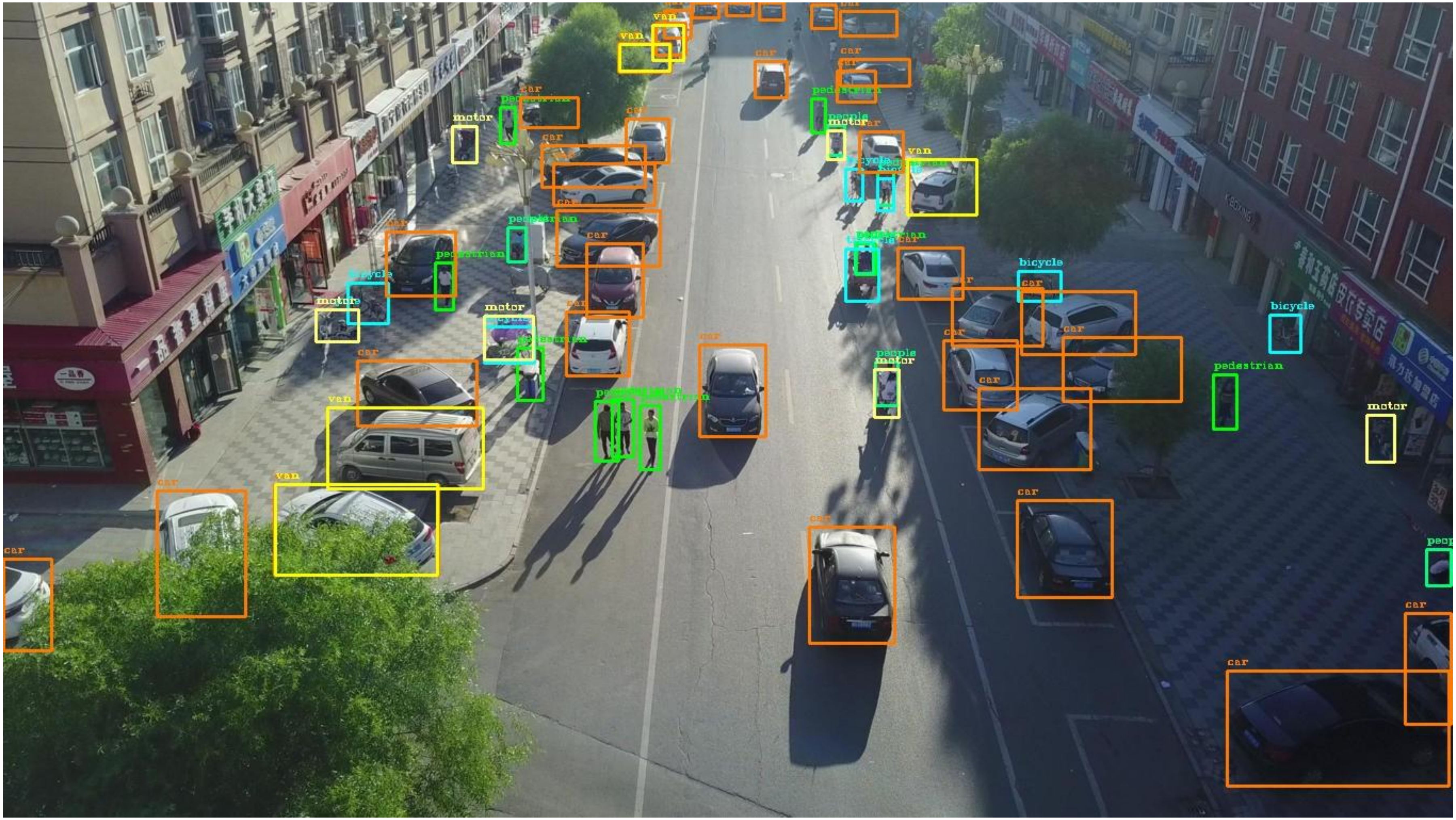
Demos


car
pedestrian
car
car
pedestrian
car
bicycle
car
car
car
motor
pedestrian
car
car
bicycle
motor
van
pedestrian
pedestrian
car
car
car
pedestrian
car
bicycle
pedestrian
car
car
car
car
car
car
motor
bicycle
car
car
car
van
motor
pedestrian
van
pedestrian
car
motor
van
motor

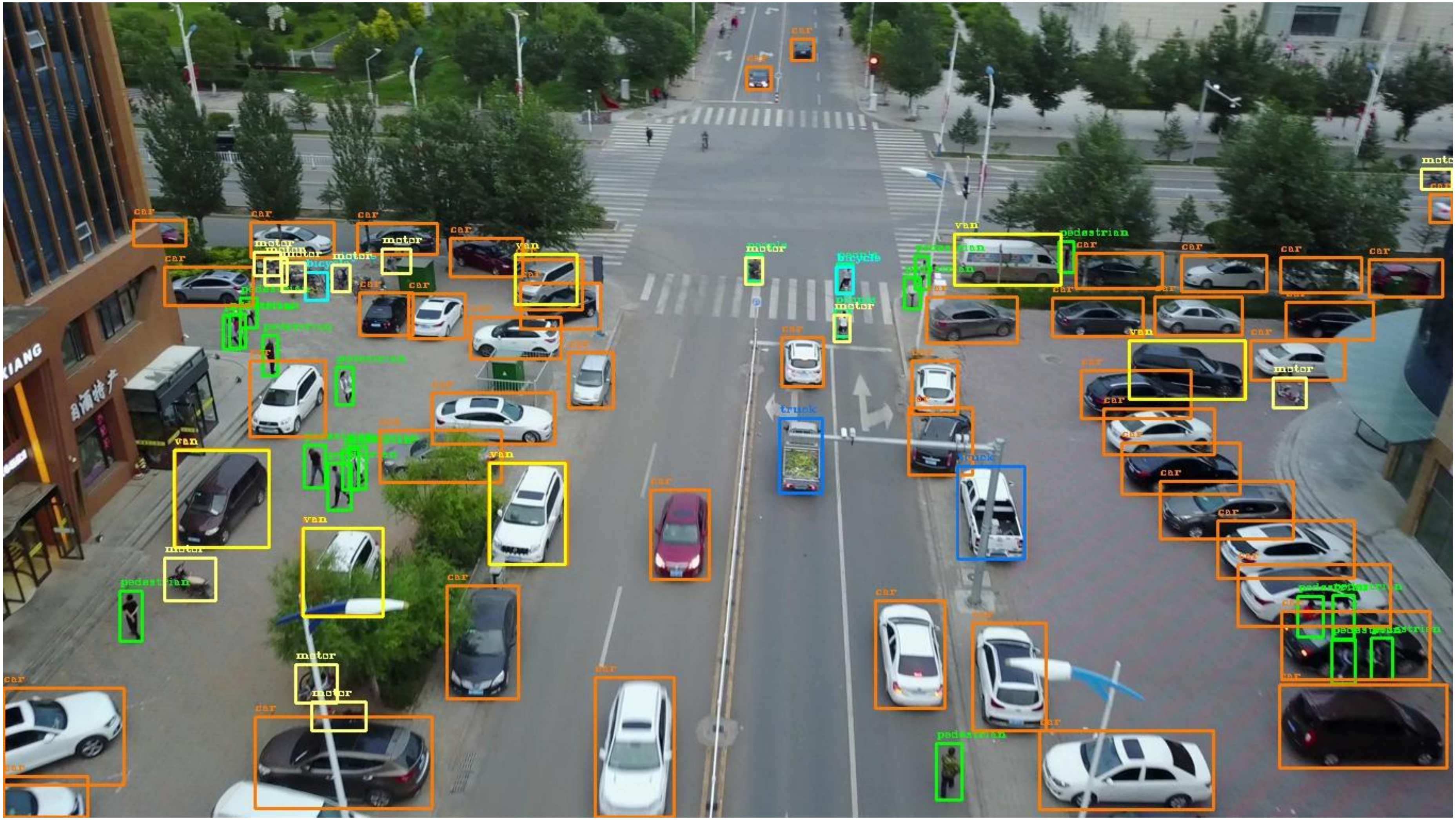
Demos


car
car
car
awning-tricycle
car
car
car
car
car
truck
van
pedestrian
pedestrian
car
pedestrian
car
car
motor
motor
motor
motor
pedestrian
bicycle
pedestrian
car
car
motor
pedestrian
motor
motor
motor
motor
motor
motor
motor
motor
motor
motor
motor
awning-tricycle
car
car
bicycle
pedestrian
bicycle
pedestrian
pedestrian
truck
van
van
car
pedestrian
张北最大的室内亲子乐园
儿童乐园
中国电信
好又多2000平米地下室火热招商中

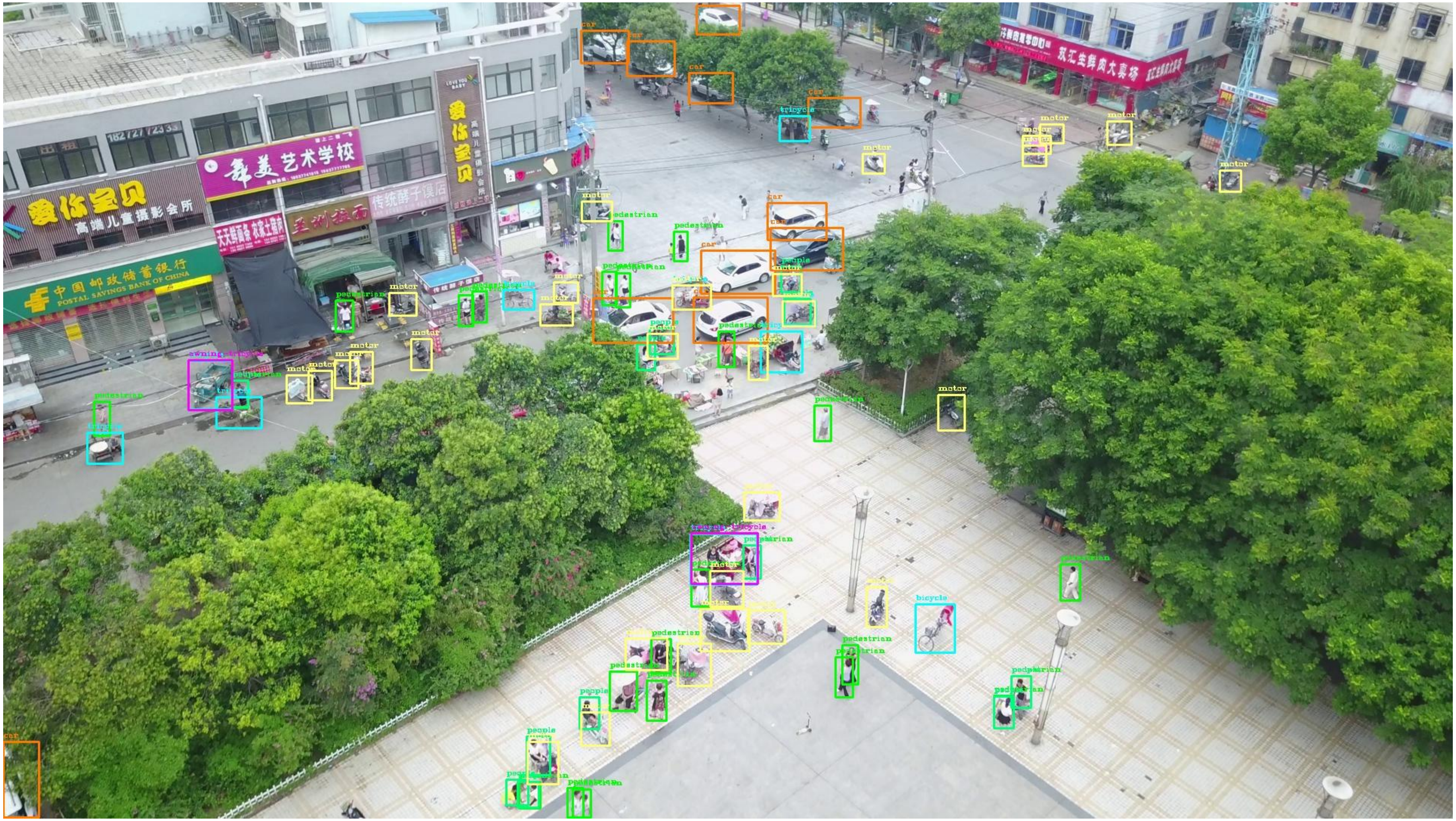
Demos


van
van
car
car
car
car
pedestrian
car
pedestrian
motor
car
people
motor
car
car
bicycle
van
pedestrian
car
pedestrian
car
pedestrian
car
bicycle
pedestrian
bicycle
car
motor
car
motor
bicycle
car
car
bicycle
car
car
people
motor
pedestrian
car
car
van
pedestrian
motor
car
car
van
car
car
peop
car
car

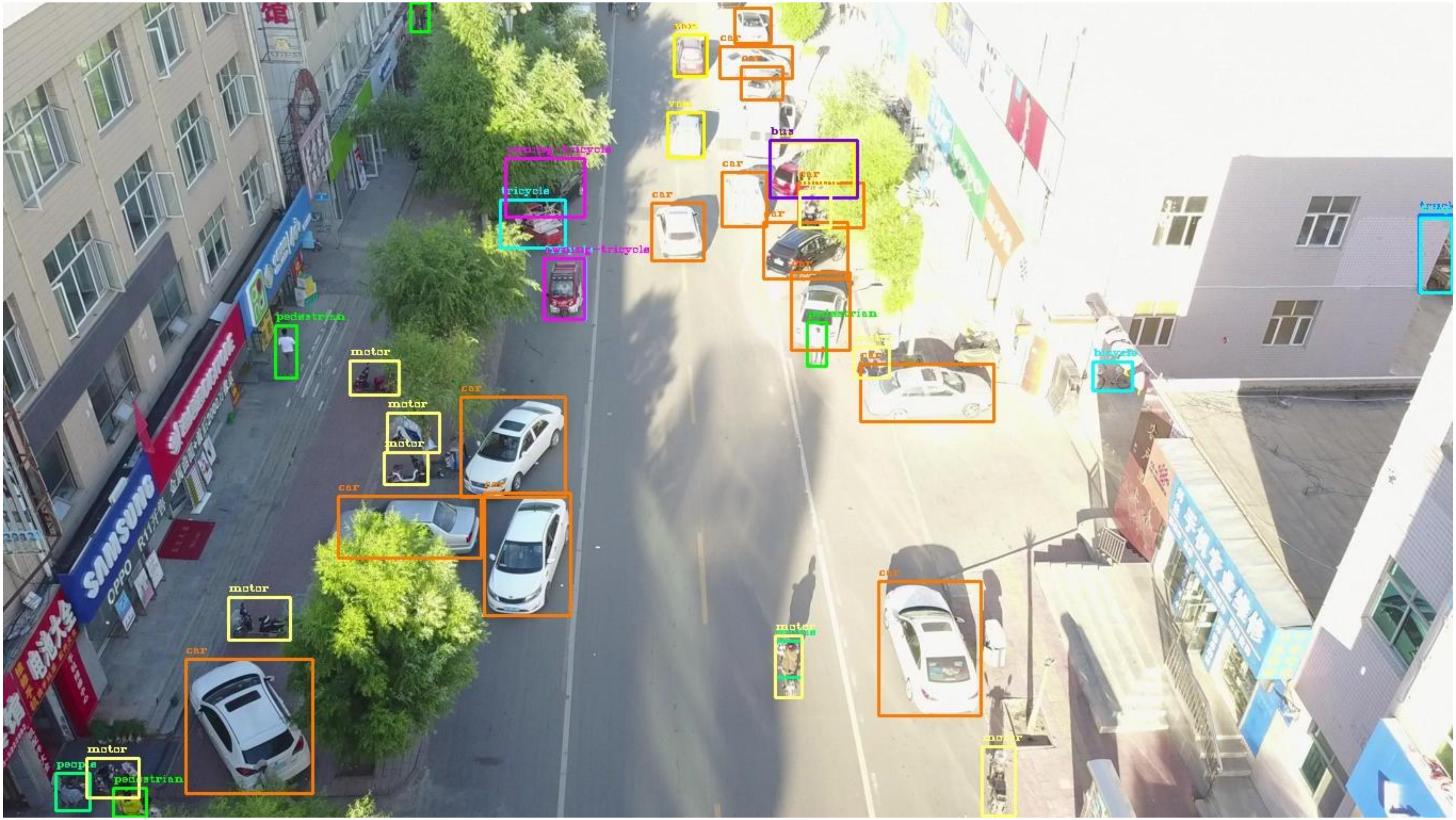
Demos



Demos


car
tricycle
motor
pedestrian
people
awning-tricycle
bicycle
舞美艺术学校
爱你宝贝
高端儿童摄影会所
中国邮政储蓄银行
POSTAL SAVINGS BANK OF CHINA
传统酵子馒店
双汇生鲜肉大卖场

Demos


van
car
car
van
bus
car
awning-tricycle
tricycle
car
car
car
awning-tricycle
car
truck
pedestrian
pedestrian
motor
car
motor
motor
car
car
bicycle
car
motor
car
motor
motor
people
pedestrian
motor

Thanks for your attention!