Google AI

OPEN IMAGES CHALLENGE 2019





Google AI

Object Detection track





Outline

Object detection track overview
Dataset
Metrics
Result analysis

Participation and winning requirements

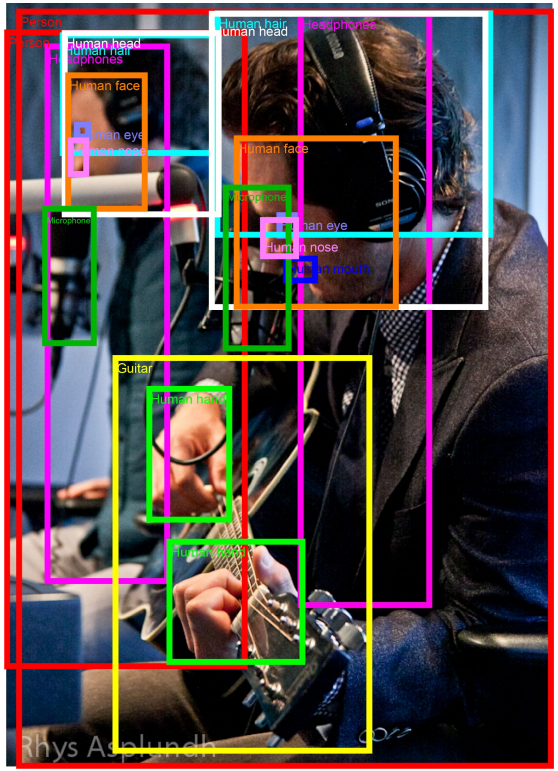
- Subset of Open Images V4/V5 used for training
- External data/pre-trained models are allowed but must be disclosed
- Evaluation server is hosted by Kaggle
- Full prize: 25K USD split between 5 winners
- Winner obligations:
    - Detailed, minimum 2-page description of method
    - Open-source model predictions
- Winners encouraged:
    - Open-source their framework

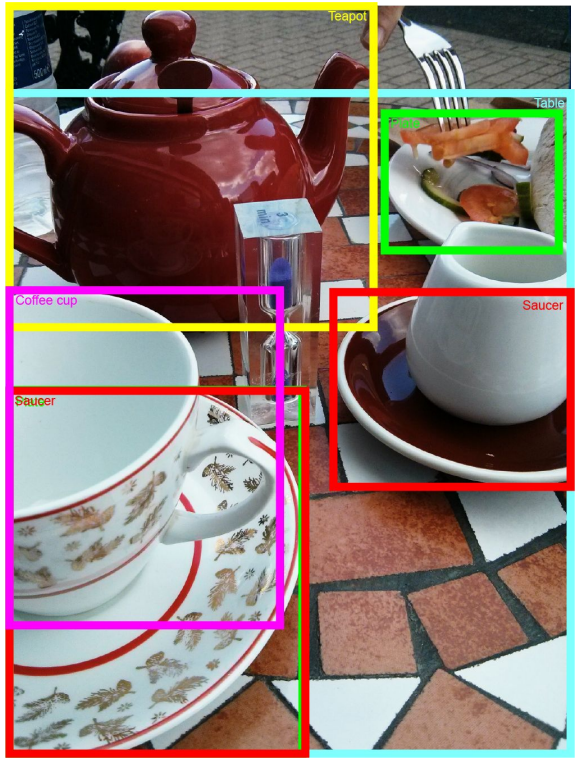
Object detection

<u>Challenging detection dataset!</u>

Key features:

- 500 annotated classes
- diverse scenes
- non-exhaustive annotation on the image-level labels level





Dataset: statistics

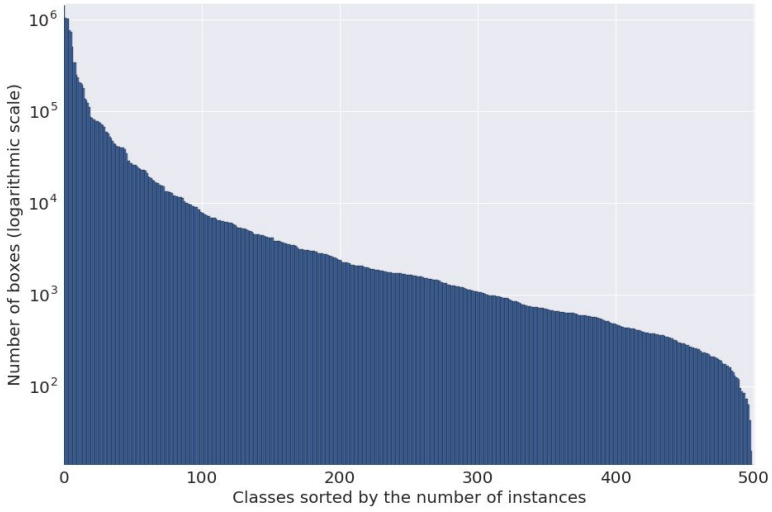
Train set:

- 1,743,042 images
- 1,913,455 negative image-level labels
- 3,830,005 positive image-level labels
- 12,195,144 boxes

Validation set of 41k images

Challenge test set:

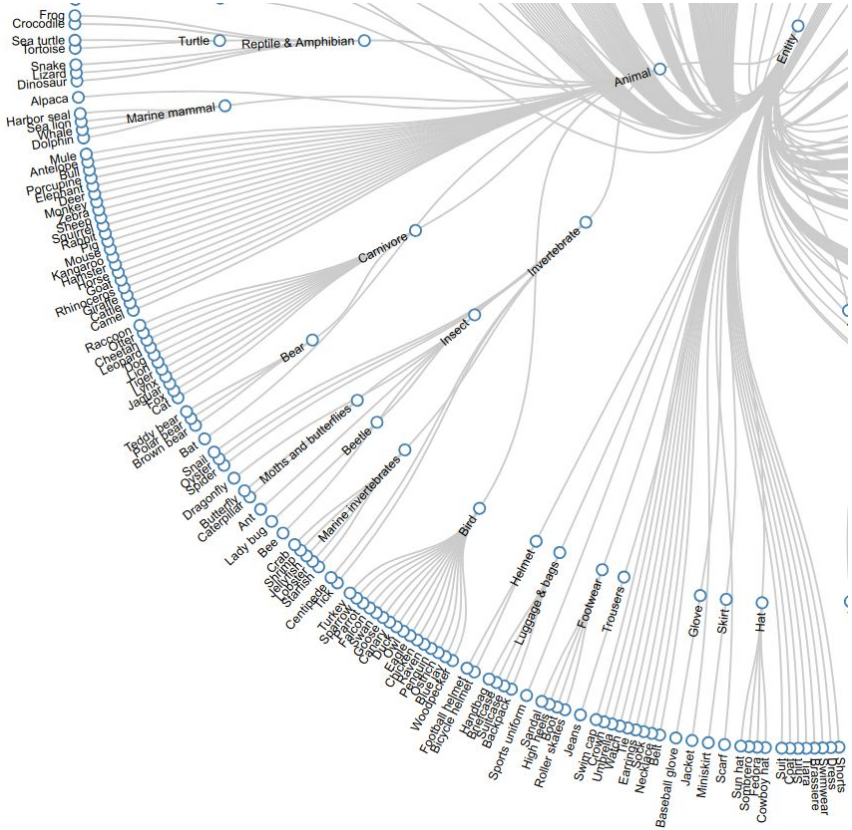
- 100K images
- 20% in public split
- 80% in private split



Dataset: class hierarchy

Classes organized in a [hierarchy]

Total: 500 classes
Leaf classes: 442 classes
Non-leaf classes: 58 classes



Evaluation

Properties of annotation process:

- Non-exhaustive image-level labeling
- Semantic hierarchy
- Group-of boxes

A modification of Mean Average Precision (mAP) takes those properties into account

Evaluation server hosted by Kaggle

Public metric implementation is available as a part of Tensorflow Object Detection API

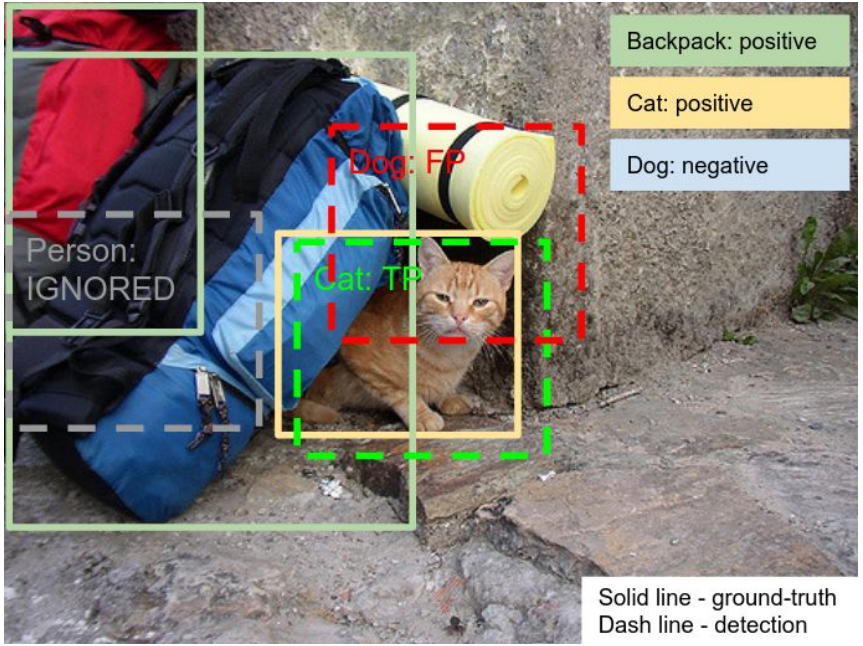
Evaluation metrics: Non-exhaustive image-level labeling

Ground-truth image labels have 3 cases:

- Positive: class is present
- Negative: class is absent
- Unannotated: we do not know

Ignore detections of unannotated classes

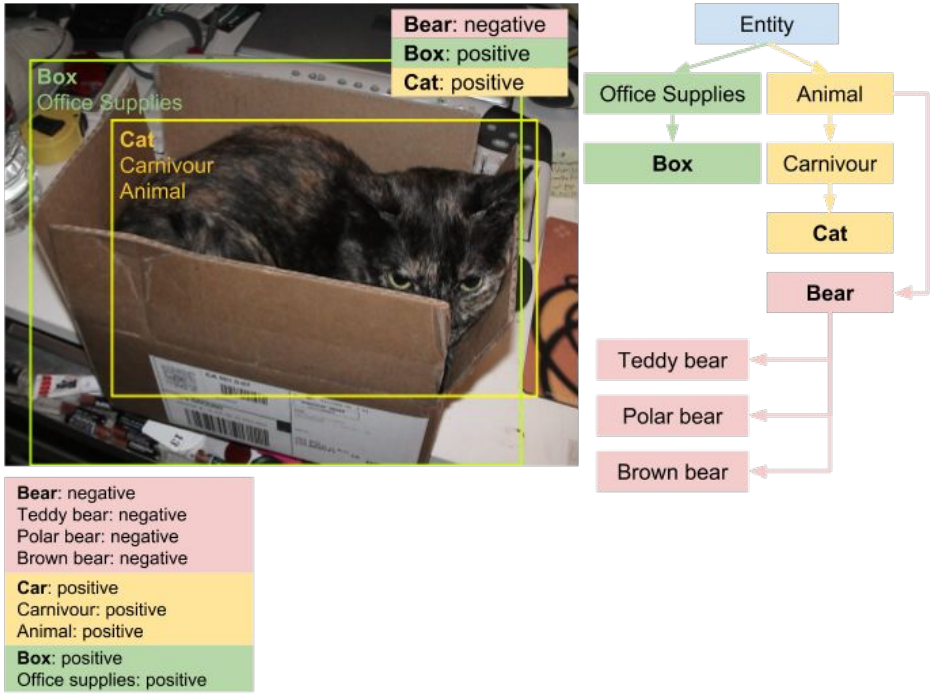
Rest as in PASCAL VOC Challenge[1]



[1]The PASCAL Visual Object Classes Challenge 2010 (VOC2010) Results, Everingham, M. and Van Gool, L. and Williams, C. K. I. and Winn, J. and Zisserman, A.

Evaluation metrics: Semantic hierarchy

- Ground truth replicates boxes and image labels following the hierarchy
- AP is computed for both leaf and non-leaf classes.
- AP for non-leaf classes is evaluated on both boxes of this class and all descendant class boxes
- Participants should output multiple boxes on same object

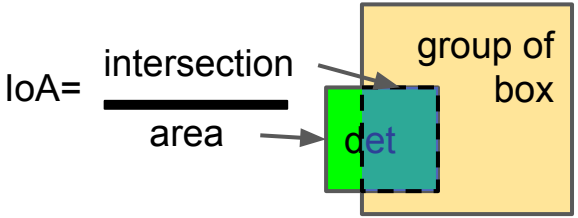


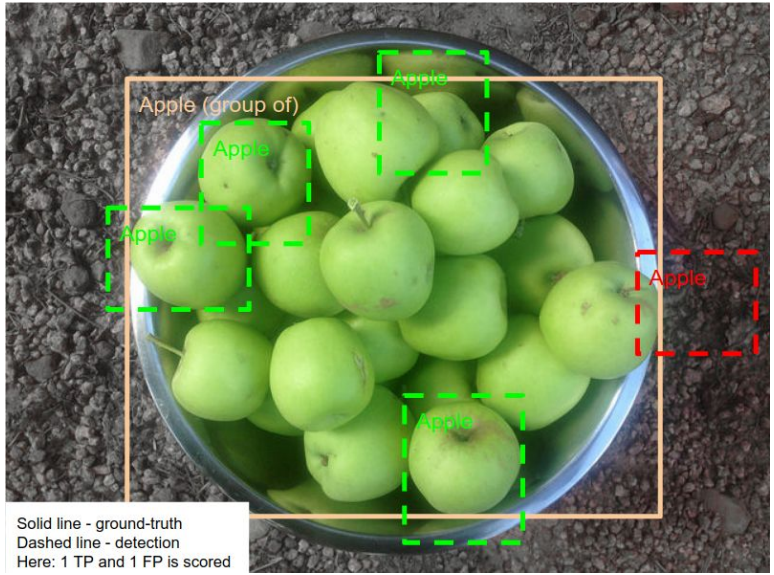
Evaluation metrics: Group-of boxes

A group-of box:

- contains >5 instances
- Instances occlude each other

Matched box:
IoA(group of box, detection) > 0.5





The highest-scoring detection is a TP. Rest is ignored

Results analysis: overview

Number of teams with at least one submission: **559**
Evaluation server is up for **4 month**

External datasets/pre-trained models used:

- Objects365
- COCO
- ImageNet
- LVIS
- ..

Base model architectures:

- FasterRCNN

Deep learning frameworks:

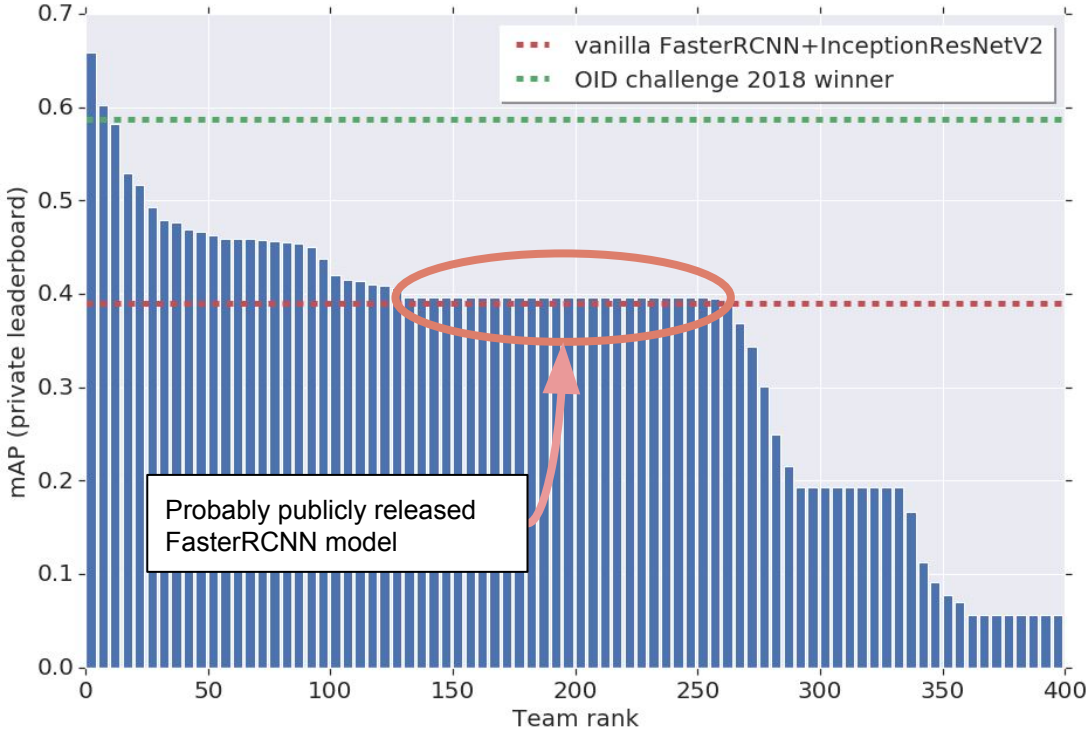
- mmdetection
- PaddlePaddle
- Chainer
- Maskrcnn-benchmark
- tf-hub models
- ...

Results analysis: teams

Number of teams: **559**
Faster-RCNN baseline: **0.39**

Number of teams beating baseline: **128**

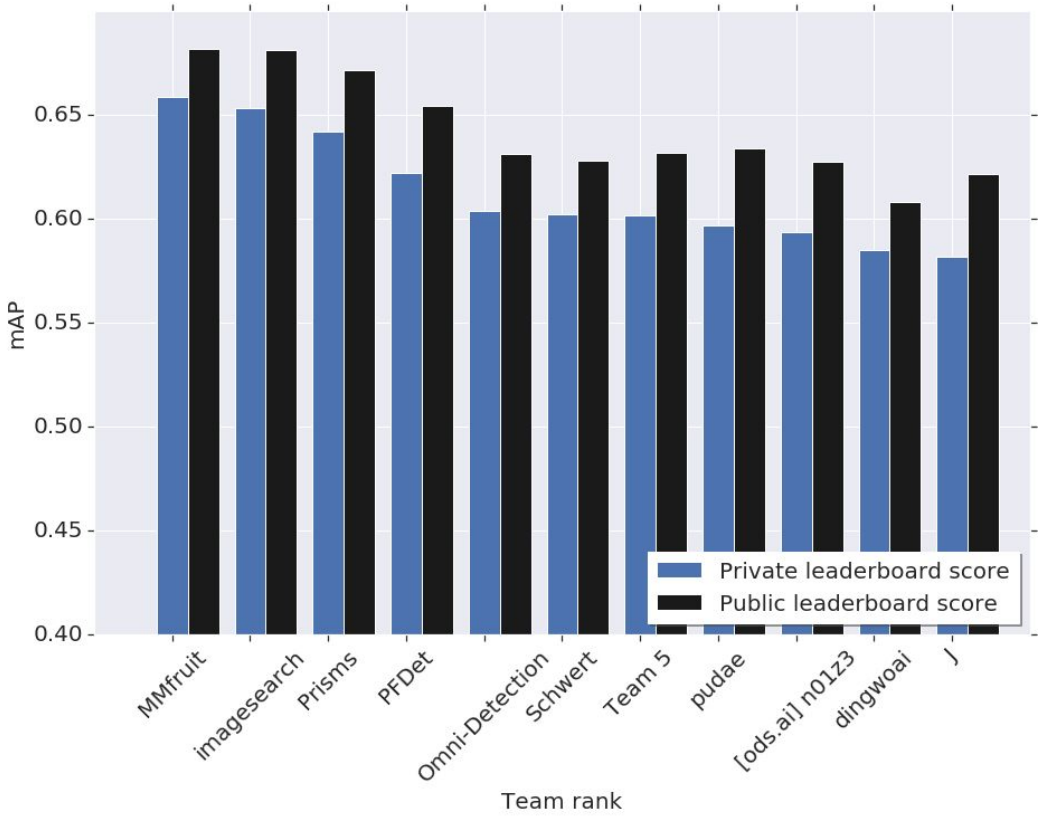
Best score of the previous year: **0.59**
Best score this year: **0.65887**



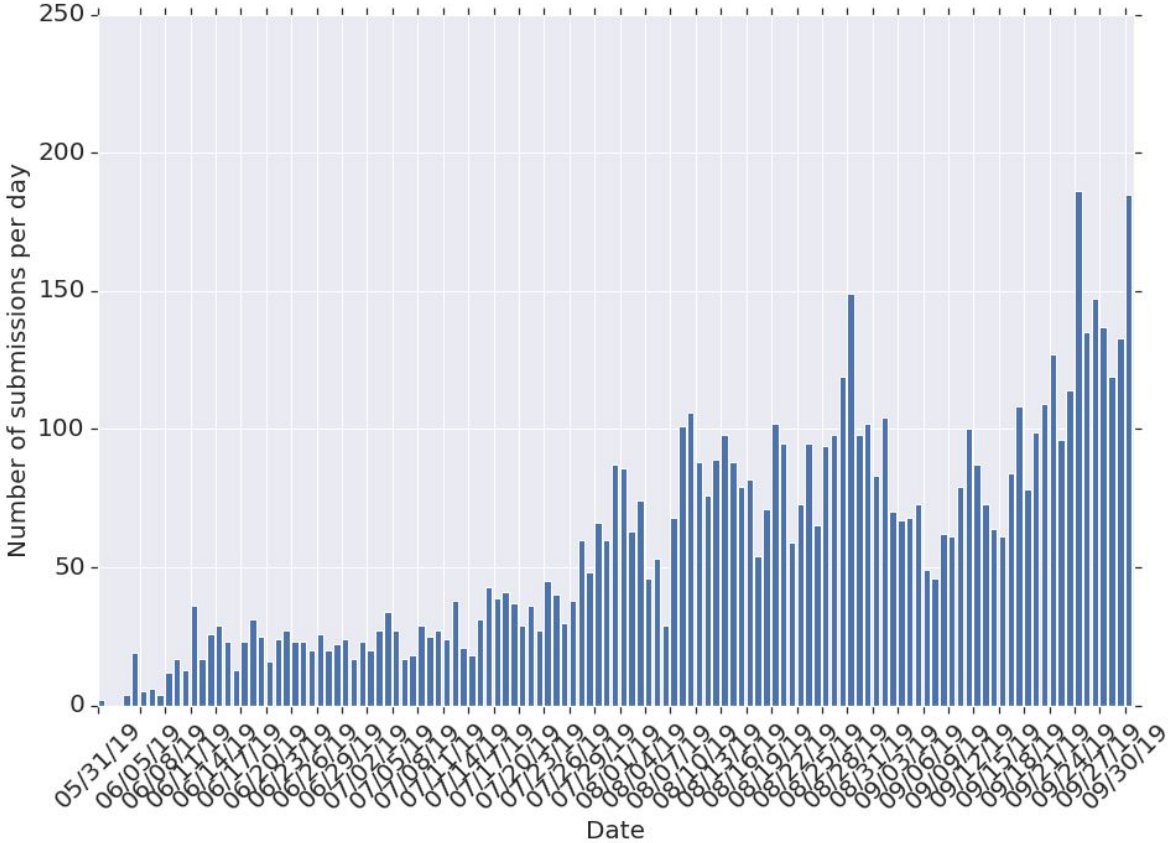
Results analysis: public vs private leaderboards

Public leaderboard: 20%
Private leaderboard: 80%

Data is identically distributed

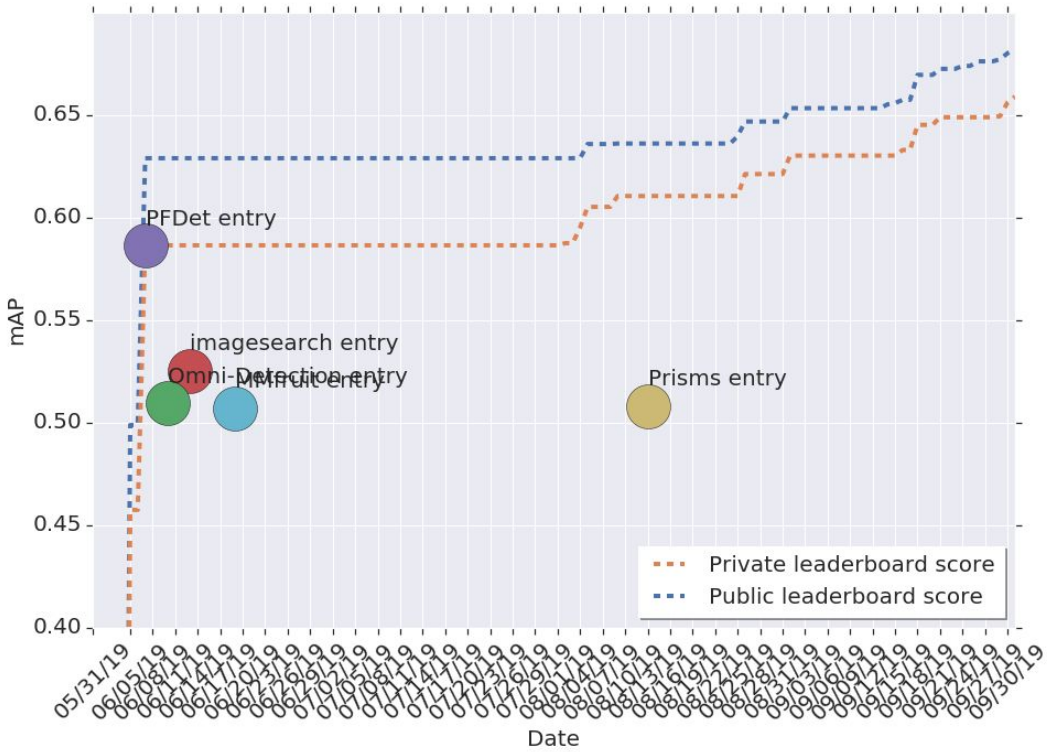


Results analysis: number of submissions per day

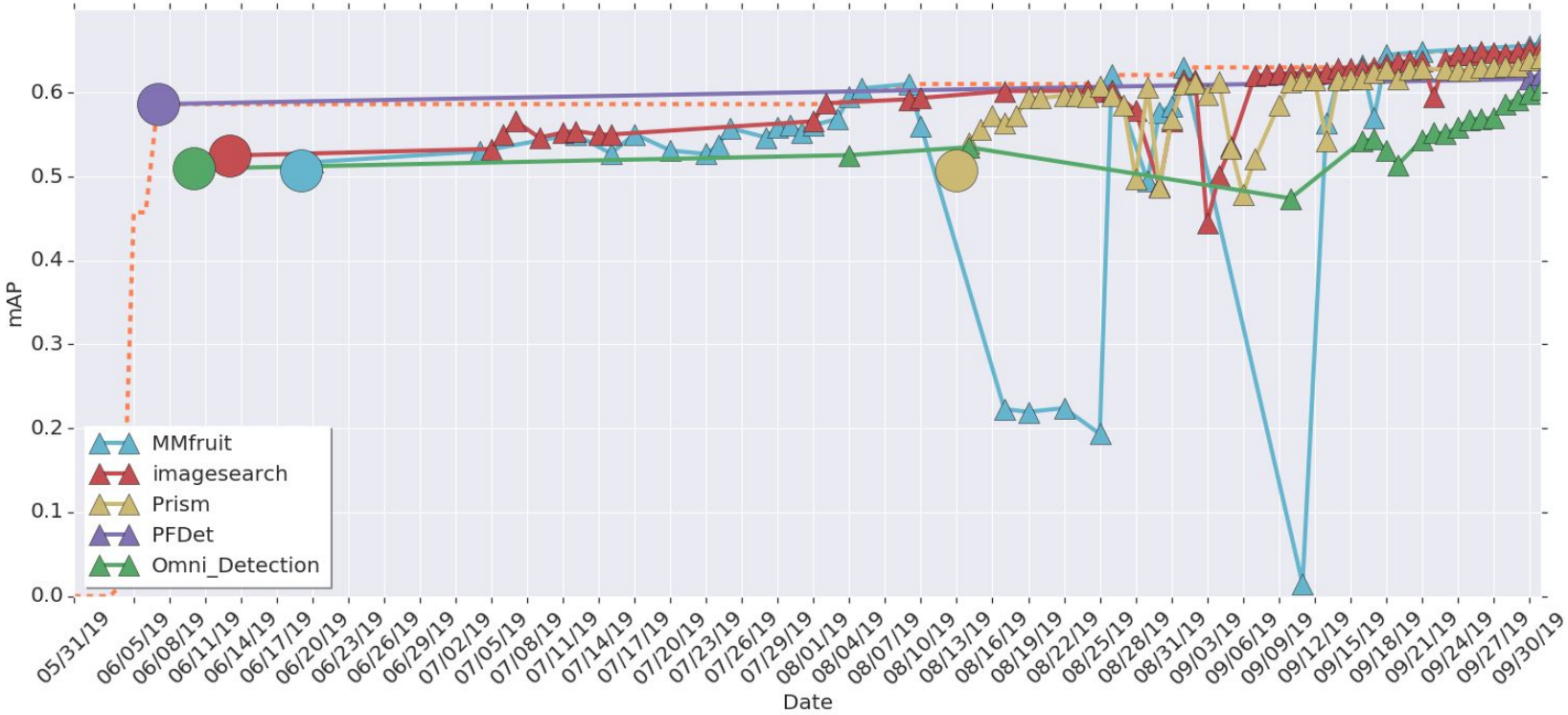


Results analysis: evolution of maximal leaderboard score

Dots: winners entering the competition



Results analysis: evolution of scores (winning teams)



Winning teams: final results

Team	Public score	Private score
MMfruit	0.68174	0.65887
imagesearch	0.68155	0.65337
Prisms	0.67170	0.64214
PFDet	0.65454	0.62221
Omni-detection	0.63408	0.60406

- First 4 teams ranked the same on private vs public leaderboard
- 5th team ranked 7th on public leaderboard

Winning models

Commonalities:

- Model ensembles
- Specialist models to handle class imbalance
- Modification of NMS (soft NMS, voting, etc)
- Several teams used image-level labels in training

Questions?

Next - presentations by winning teams