Google AI

OPEN IMAGES

CHALLENGE 2018

Google AI

Visual Relationship Detection track

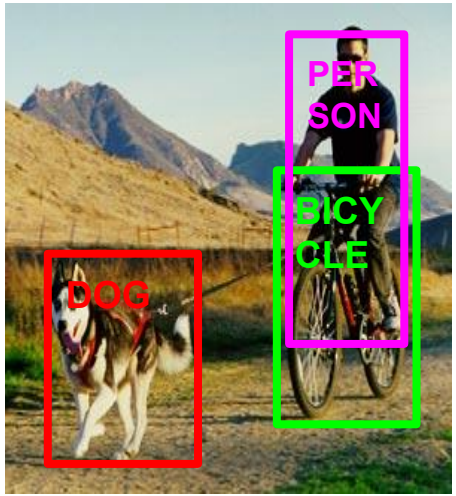
Outline

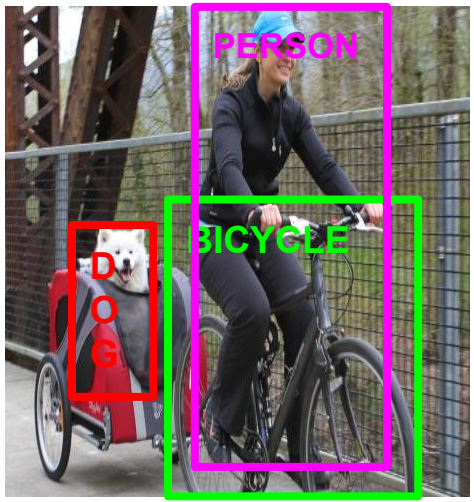
- Visual relationship detection track overview
- Dataset: data collection and statistics
- Metrics
- Result analysis

Visual relationship detection

Task:

- Two objects locations and classes
- Relationship between two objects





Both images have the same set of objects and layout but very different semantics

Participation and winning requirements

- Additional annotations on top of Open Images V4
- External data/pre-trained models are allowed but must be disclosed
- Evaluation server is hosted by Kaggle
- Full prize: 20K USD split between 3 winners
- Winner obligations:
    - Detailed, minimum 2-page description of method
- Winners encouraged:
    - Open-source their framework

Dataset: data collection

1. Existing works
   - VRD dataset[1]
   - Visual Genome Dataset[2]

2. Generate label co-occurrence statistics of Open Images V4
+ pick interesting relationships

3. Generate candidate triplets for annotation

Relationships:
<Pets> **under** {Table, Chair, etc ...}
<Object> **on top of** <Object>
<Object> **inside of** <Object>
<Human> **holds** <Object>
<Human> **on top of** <Object>
<Human> **hits** {Football,Tennis ball,...}
<Human> **plays** {Drums, Guitar, ...}
<Object> **is** {attribute}

[1]Lu, C., Krishna, R., Bernstein, M, Fei-Fei, Li, "Visual Relationship Detection with Language Priors", ECCV 2016
[2] Krishna R., Zhu Y., Groth O., Johnson J., Hata K., Kravitz J., Chen S., Kalantidis Y., Jia-Li L., Ayman Shamma D., Bernstein M., Fei-Fei L., "Visual Genome: Connecting Language and Vision Using Crowdsourced Dense Image Annotations", 2016

Dataset: annotation

Example triplet: Man **holds** Microphone

Dataset: annotation

Example triplet: Man **holds** Microphone

Dataset: annotation

Please verify that the relation **holds** connects the man and the microphone on the image: man **holds** microphone

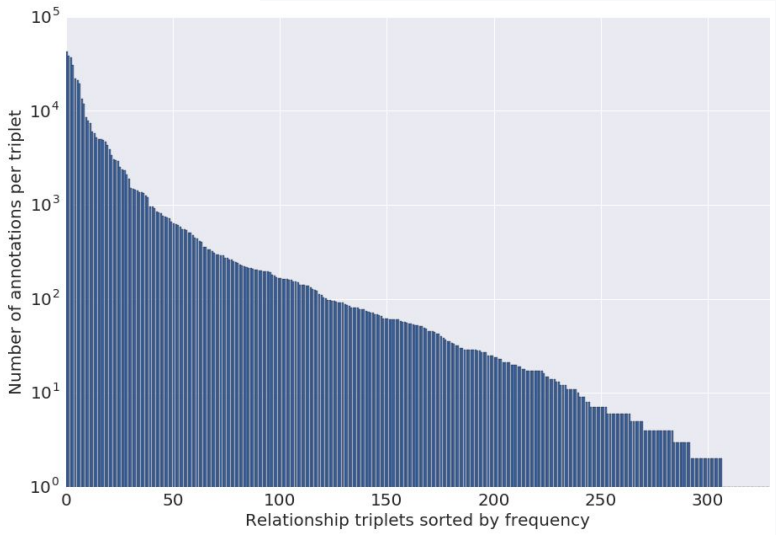
Dataset: statistics

Train set:

- 1,743,042 images
- 374,768 relationship annotations
- 3,290,070 bounding boxes
- 329 distinct triplets
- 100k subset for validation

Test set:

- 100K images
- 30% in public split
- 70% in private split



Evaluation

No standard metric for visual relationships detection evaluation.

Evaluation server is hosted by Kaggle

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

Evaluation: metrics

Three metrics used in literature[1,2]:

- AP relationships detection (but reported values are low)
- AP phrase detection
- Recall@50, Recall@100 for both relationship detection and phrase detection

Final score:

0.4*mAP(relationships) + 0.4*mAP(phrase) + 0.2*Recall@50(relationships)

[1]Lu, C., Krishna, R., Bernstein, M, Fei-Fei, Li, “Visual Relationship Detection with Language Priors”, ECCV 2016

[2] Krishna R., Zhu Y., Groth O., Johnson J., Hata K., Kravitz J., Chen S., Kalantidis Y., Jia-Li L., Ayman Shamma D., Bernstein M., Fei-Fei L., “Visual Genome: Connecting Language and Vision Using Crowdsourced Dense Image Annotations”, 2016

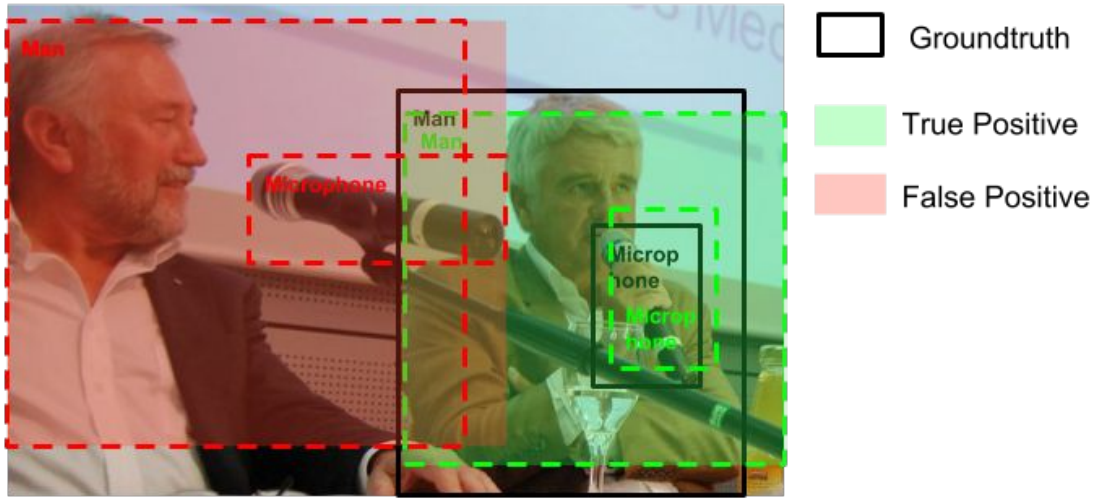
Evaluation: metrics

AP per relationship
(i.e. **holds**)

- mean AP(relationships)
- Recall@50

True Positive:

- IoU > 0.5 for each box
- Object labels and relationship label match



Evaluation: metrics

AP per relationship
(i.e. **holds**)

mean AP(phrase)

True Positive:

- IoU > 0.5 for box union
- Object labels and relationship label match

Results analysis: overview

Number of teams with at least one submission: **232 teams**
Evaluation server days: **51**

External datasets/pre-trained models used:

- OpenImagesV4
- ImageNet
- COCO
- Visual Genome

Base model architectures:

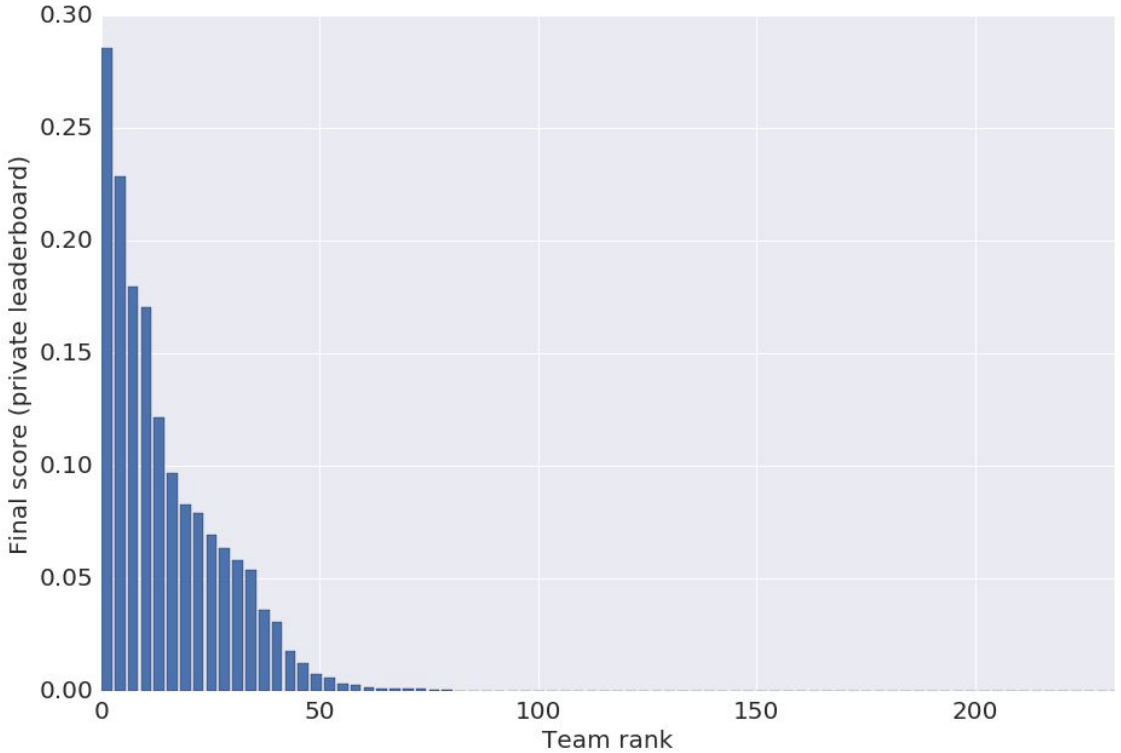
- ResNets, YOLO, Darknet, SENet, Retinanet ...

Deep learning frameworks:

- Tensorflow Object Detection API, Detectron, Cadene (pyTorch), fastai library, ImageAI, ChainerCV, TensorFlow-Slim, Keras, MXNet

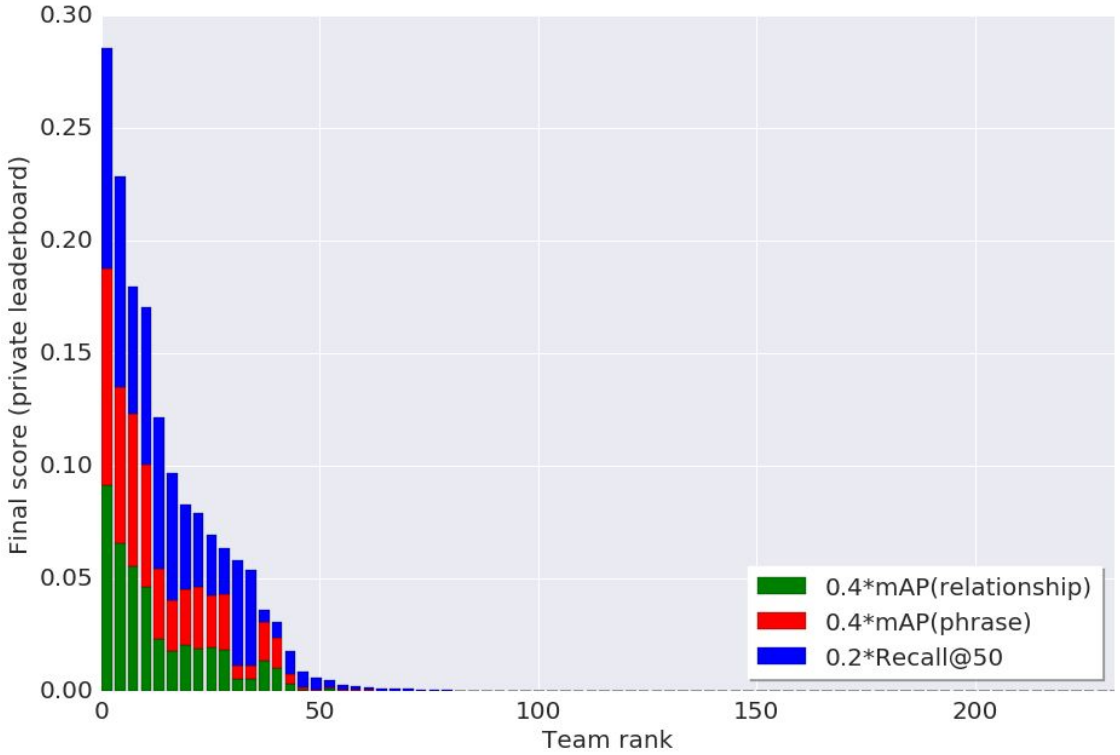
Results analysis: teams

Number of teams:
232

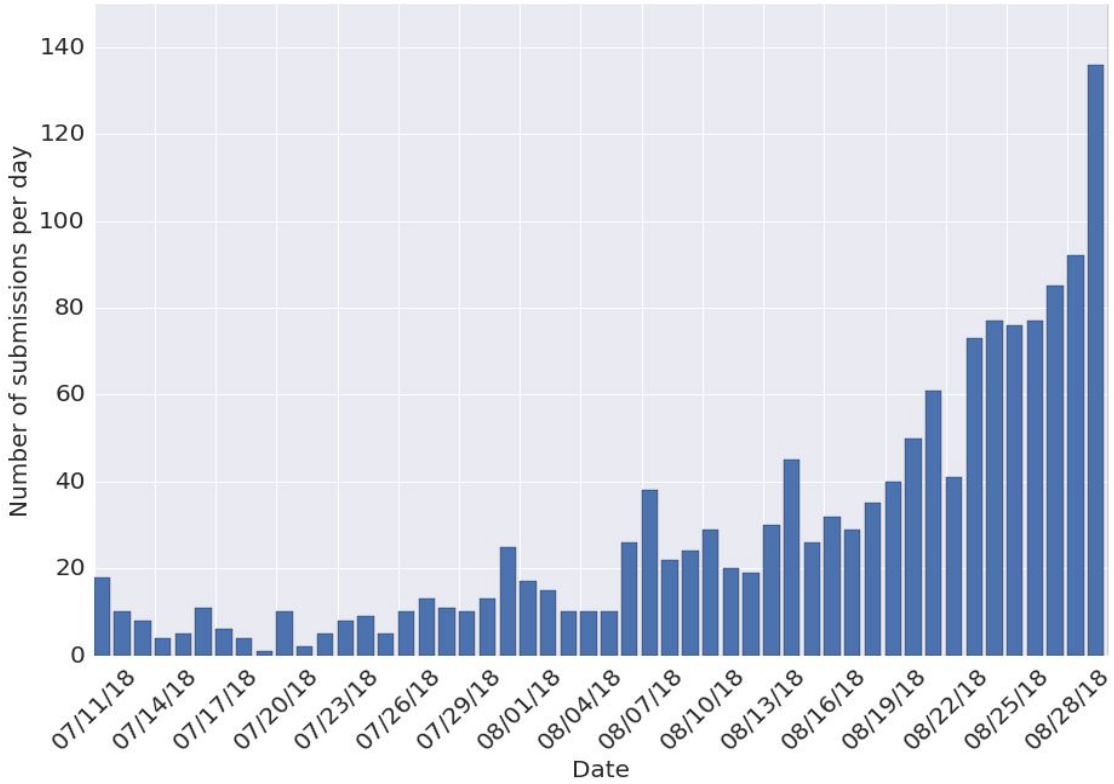


Results analysis: components of the final score (weighted)

- Some teams with lower overall score had higher score in mAPs
- Some teams scored well, mostly because of Recall@50

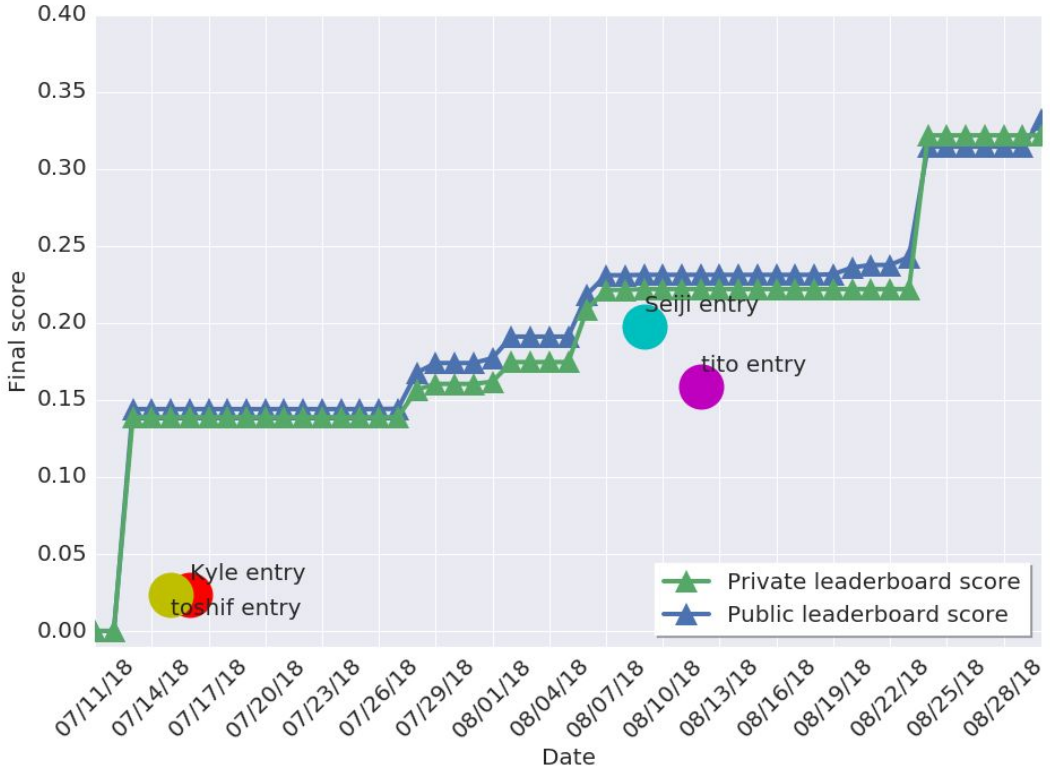


Results analysis: number of submissions per day

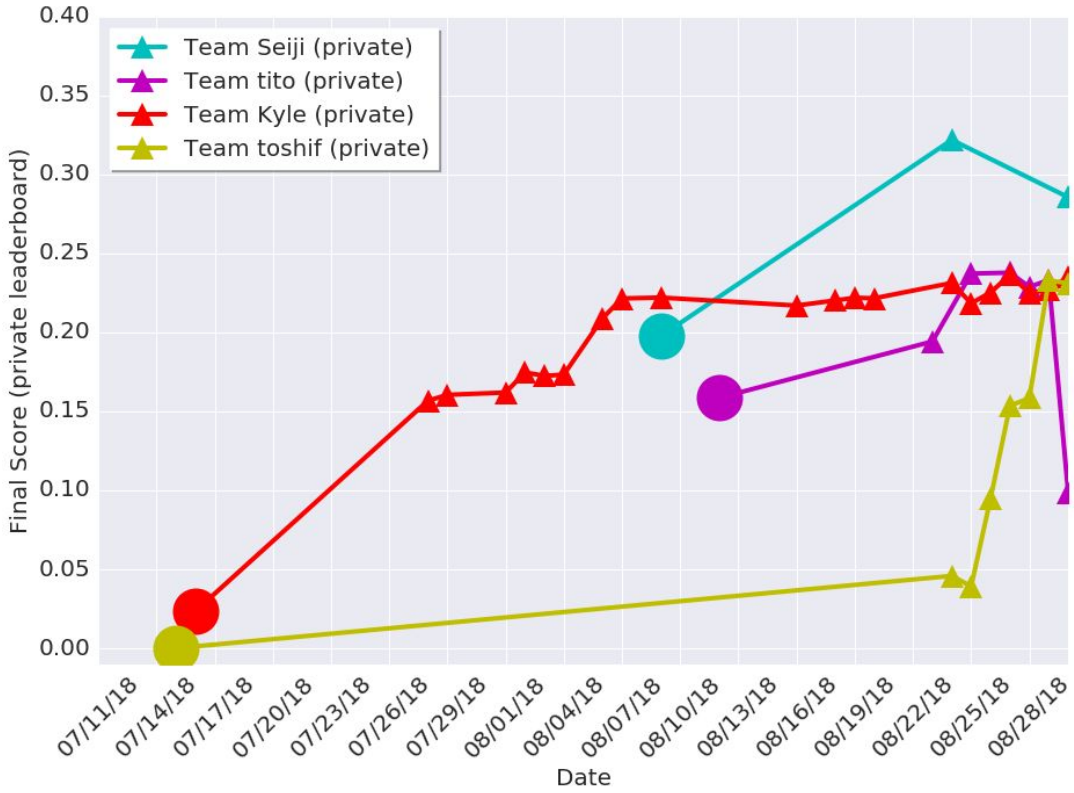


Results analysis: evolution of scores

Dots: winners entering the competition

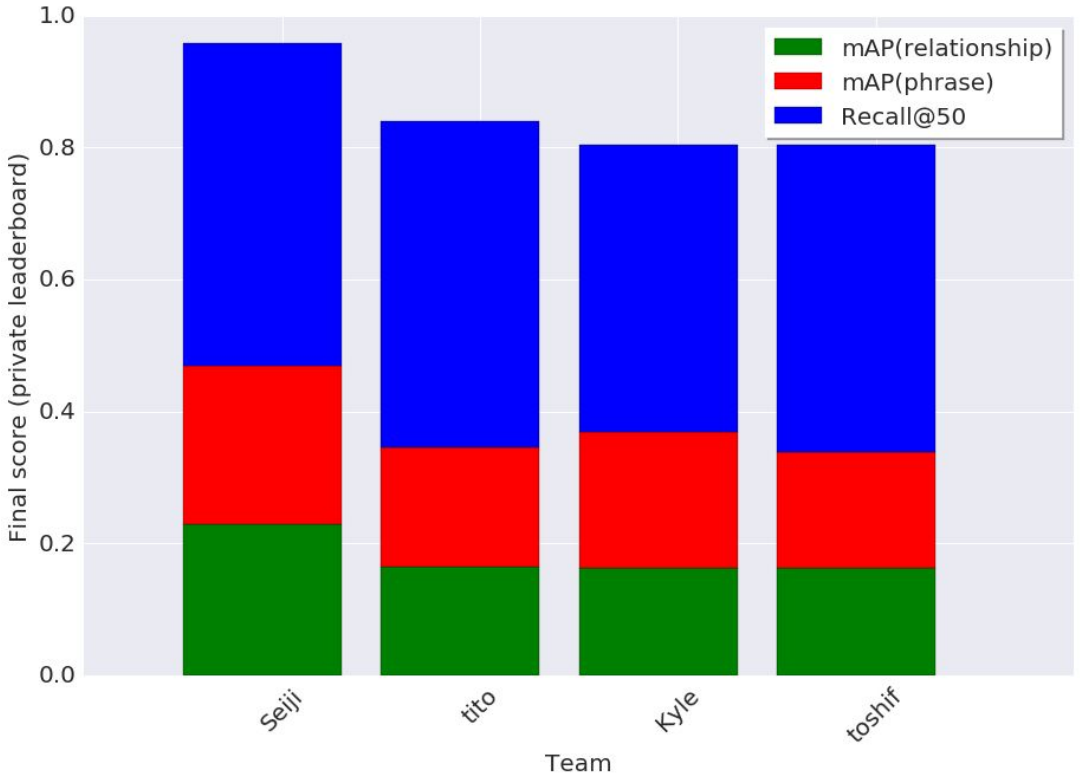


Results analysis: evolution of scores (winning teams)



Results analysis: winners breakdown by score compos (unweighted)

2rd, 3rd and 4th place teams showed approximately the same performance on mAP(relationship)



Winning models: final result

	Public leaderboard score	Private leaderboard score
Seiji	0.33213	**0.28544**
tito	0.25571	0.23709
Kyle	0.28043	0.23491
toshif	0.25621	0.22832

Undisclosed

Winning models

Commonalities:

- Different models for attributes (“is” relationship) and relationships between two objects
- The models combine a detector with a module on top for relationship prediction

Questions?

Next - presentations by winning teams

Materials below this slide

Candidates generation

Example triplet: Man **holds** Microphone

Candidates generation

Example triplet: Man **holds** Microphone

Annotation process

Please verify that the relation **holds** connects the man and the microphone on the image: man **holds** microphone

More about annotation process:

go/oi_triplet_annotation

Annotation process

Please verify that the relation **holds** connects the man and the microphone on the image: man **holds** microphone

More about annotation process:

go/oi_triplet_annotation

Annotation process

Please verify that the relation **holds** connects the man and the microphone on the image: man **holds** microphone

More about annotation process:

go/oi_triplet_annotation