Google AI

OPEN IMAGES CHALLENGE 2019

Google AI

Visual Relationship Detection track

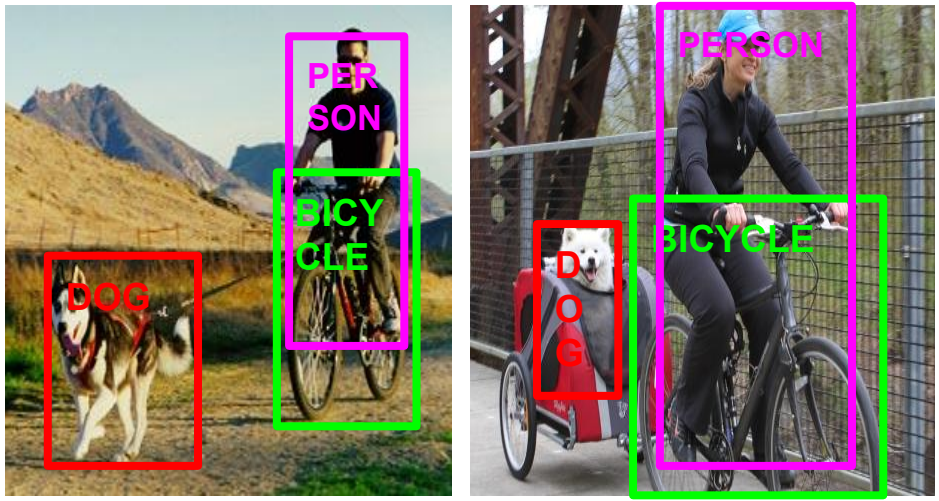
Outline

- Visual relationship detection track overview
- Dataset: statistics and metrics
- Result analysis and comparison to the previous year

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

- 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

Dataset: statistics

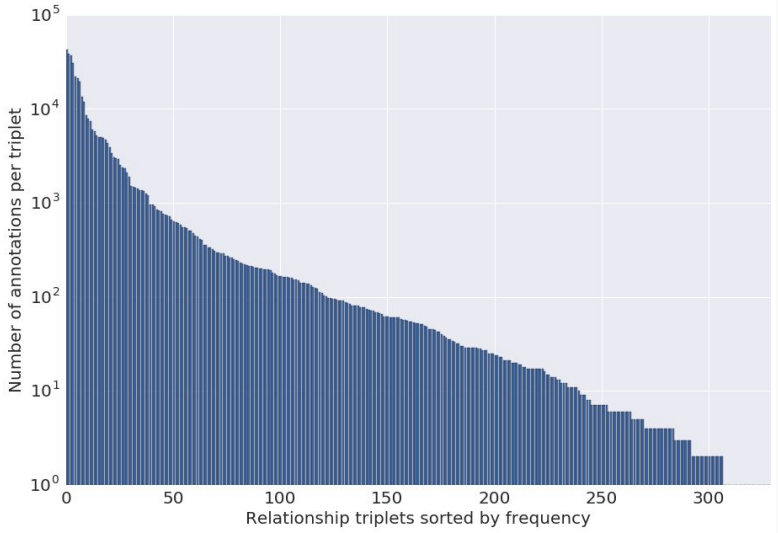
Train set:

- 1,743,042 images
- 374,768 relationship annotations
- 3,290,070 bounding boxes
- 329 distinct triplets

Validation set of 41k images

Challenge 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

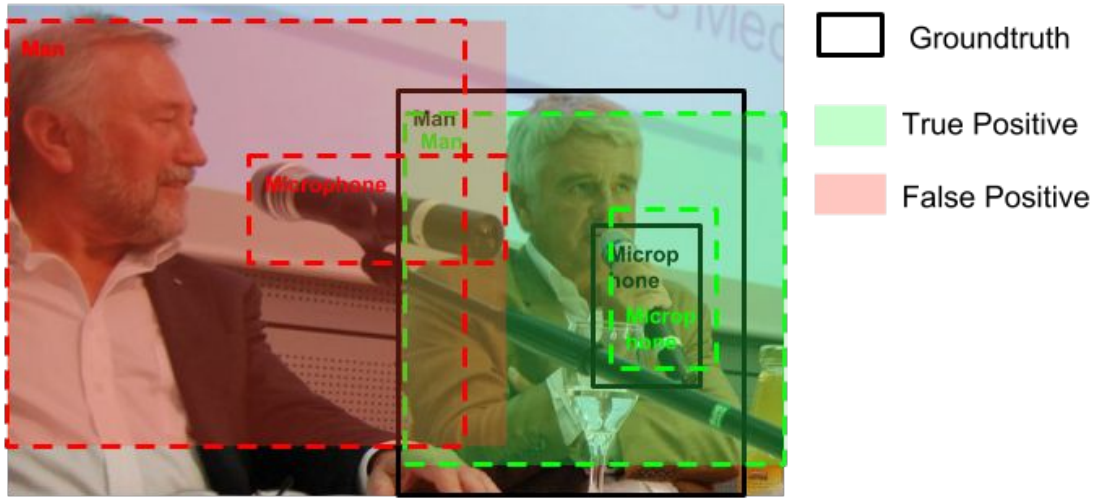
Evaluation: metrics

AP[1] per relationship
(i.e. **holds**)

- mean AP(relationships)
- Recall@50

True Positive:

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



[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

AP[1] per relationship
(i.e. **holds**)

mean AP(phrase)

True Positive:

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

[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.

Results analysis: overview

Number of teams with at least one submission: **201 teams**
Evaluation server is up for 4 months

External datasets/pre-trained models used:

- COCO
- Objects365
- ImageNet
- ...

Base model architectures:

- FasterRCNN
- YOLO
- ...

Deep learning frameworks:

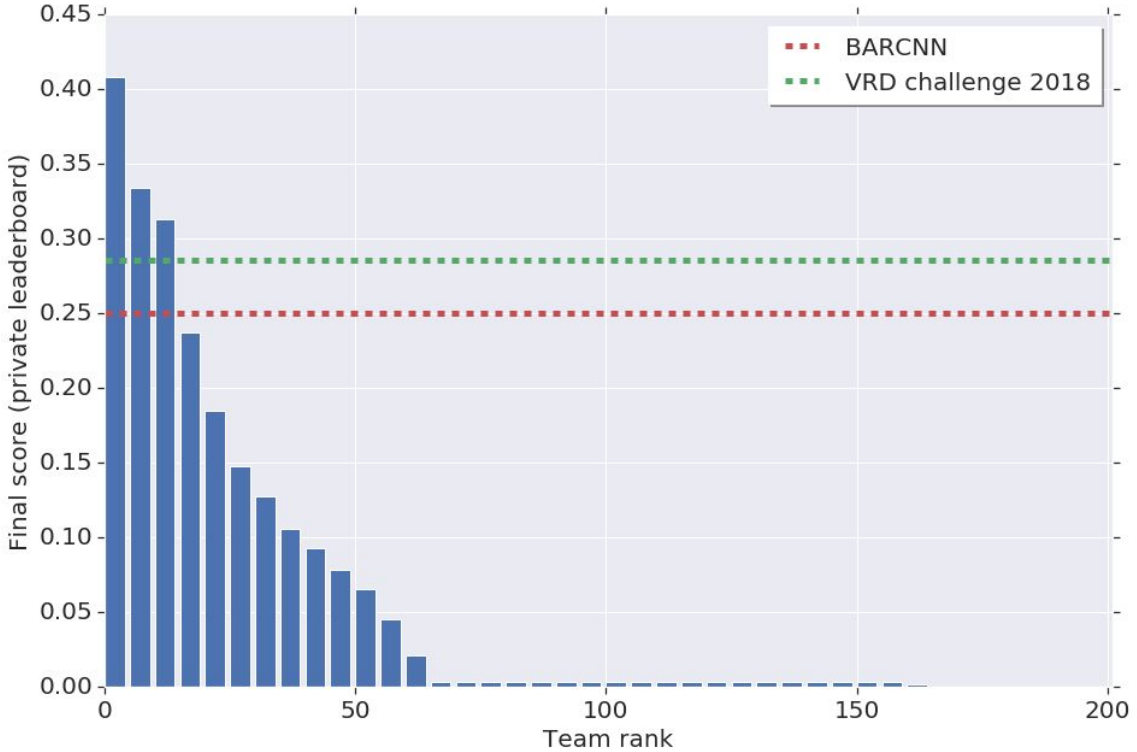
- mmdetection
- tf-hub modules
- Tensorpack
- ...

Results analysis: teams

Number of teams: **201**
BAR-CNN baseline[1]: **0.25**

Number of teams beating the baseline: **14**

Previous year best result: **0.28**
This year result: **0.40801**

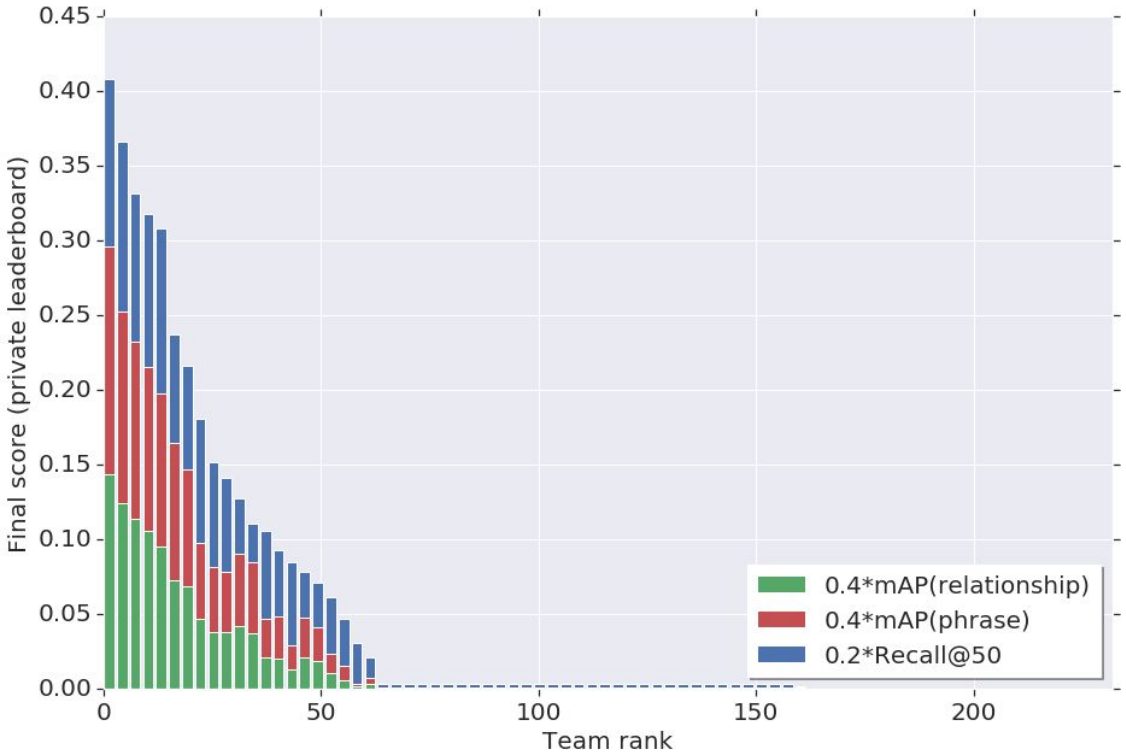


[1] Detecting Visual Relationships Using Box Attention, A. Kolesnikov, A. Kuznetsova, C. H. Lampert, V. Ferrari, ICCV Workshops, 2019

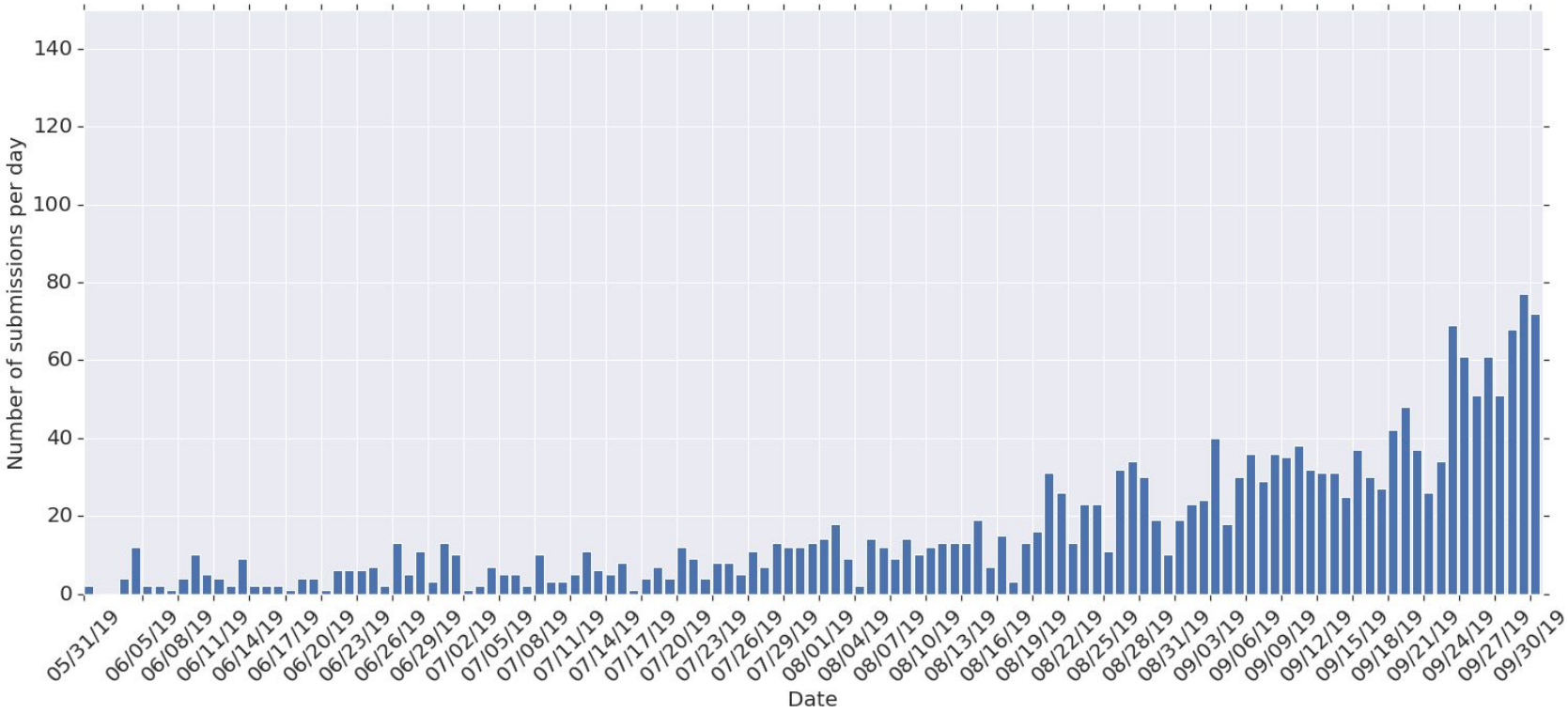
Results analysis: components of the final score (weighted)

mAP for relationship detection is still quite low (~0.36)

Some teams have high recall, but other scores are very low



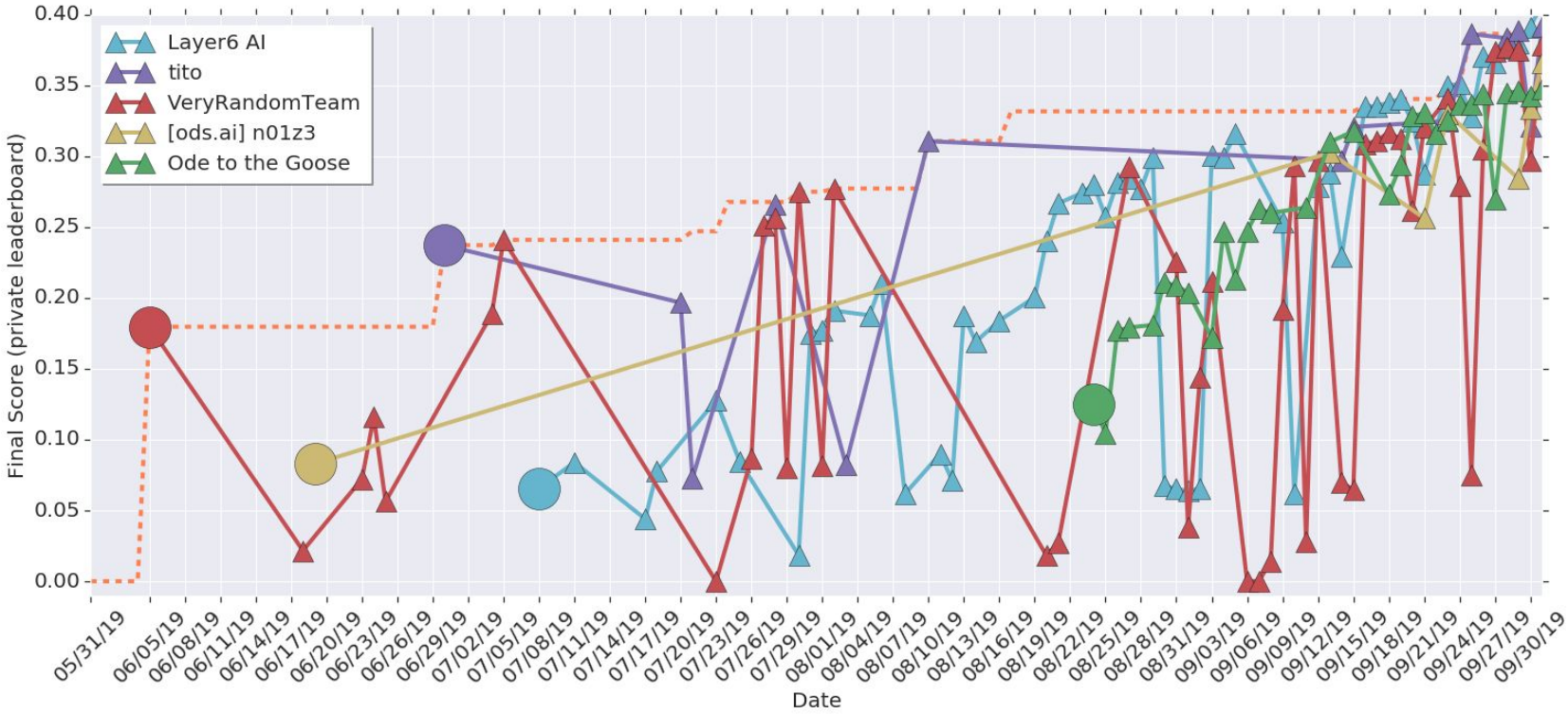
Results analysis: number of submissions per day



Results analysis: evolution of scores

Dots: winners entering the competition

Results analysis: evolution of scores (winning teams)



Winning models: final result

	Public score	Private score
Layer6 AI	0.46382	0.40801
tito (2d place last year)	0.44079 (0.25571 last year)	0.38818 (0.23709 last year)
Very Random team	0.42894	0.37853
[ods.ai] n01z3	0.39847	0.36597
Ode to the Goose	0.40165	0.34779

- First 3 teams ranked the same on private vs public leaderboard
- 4th and 5th teams switched places

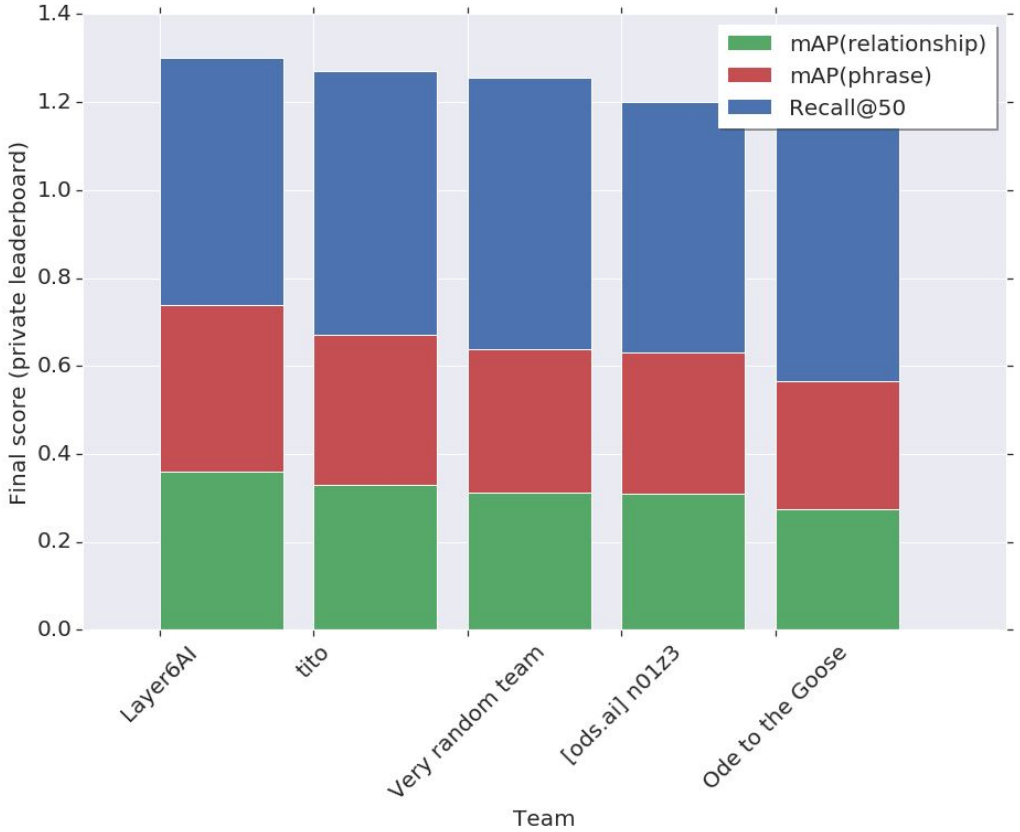
Winning models

Commonalities:

- Different models for attributes (“is” relationship) and relationships between two objects (due to performance)
- For relationships: detection model + model on top using detections
- Non-deep learned models to capture spatial and semantic information (GBMs) on top of detections (non-visual features)
- CNNs for working with visual features

Results analysis: winners breakdown by score components (unweighted)

mAP (relationship) is still low (**~0.36**) compared to last year: **~0.21**



Questions?

Next - presentations by winning teams