COCO
Common Objects in Context

1st Keypoints Challenge

Yin Cui, Tsung-Yi Lin, Matteo Ruggero Ronchi, Genevieve Patterson

ImageNet and COCO Visual Recognition Challenges Workshop

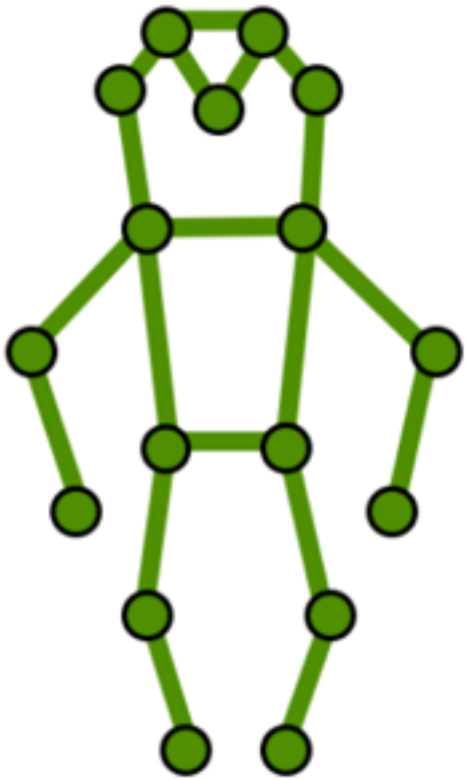
Sunday, October 9th, ECCV 2016

COCO
Common Objects in Context

NEW Keypoints Dataset

Dataset Statistics (I)

Keypoints Dataset

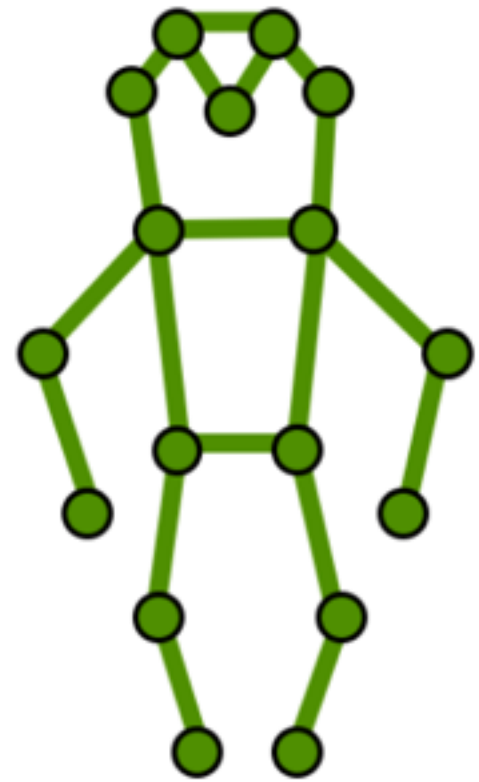


Dataset Statistics (I)

COCO
Common Objects in Context

Keypoints Dataset

- 17 types of keypoints.

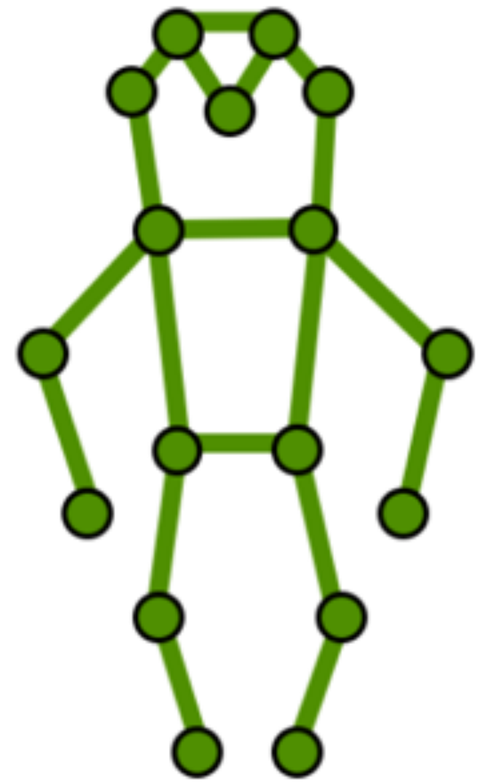


Dataset Statistics (I)

COCO
Common Objects in Context

Keypoints Dataset

- 17 types of keypoints.
- Visibility flag annotations.

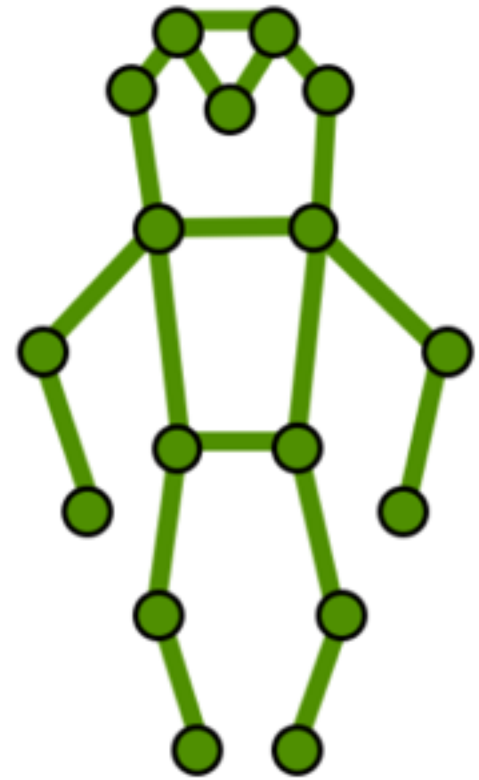


Dataset Statistics (I)

COCO
Common Objects in Context

Keypoints Dataset

- 17 types of keypoints.
- Visibility flag annotations.
- 105,698 annotated people.

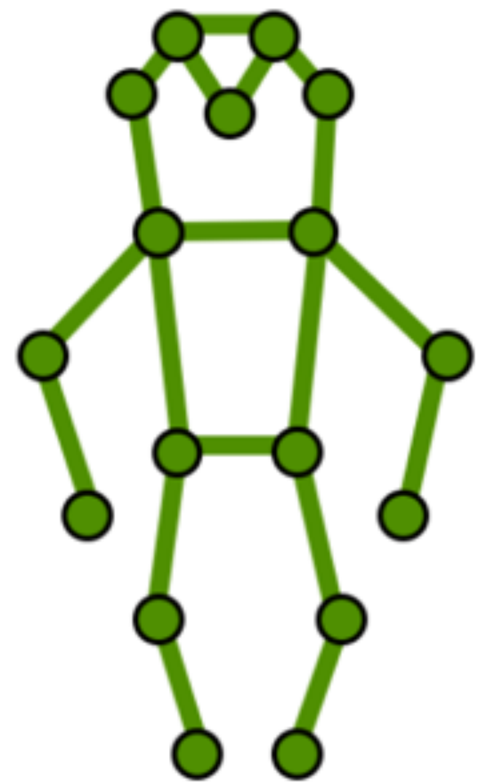


Dataset Statistics (I)

COCO
Common Objects in Context

Keypoints Dataset

- 17 types of keypoints.
- Visibility flag annotations.
- 105,698 annotated people.
- 1,161,667 annotated keypoints.

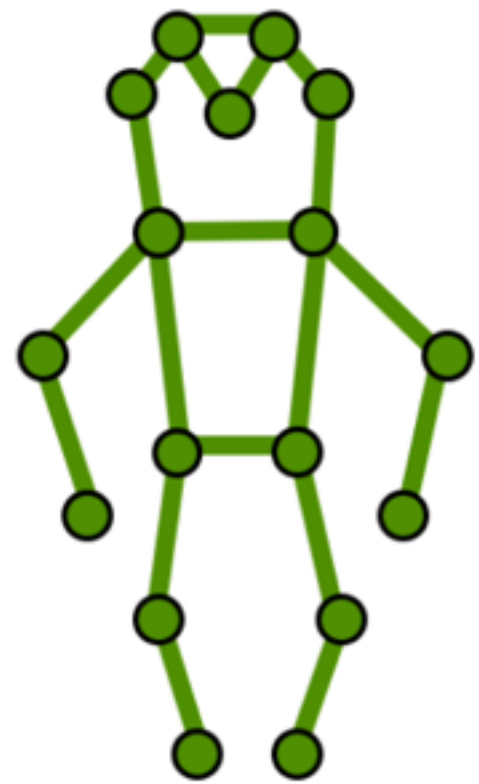


Dataset Statistics (I)

COCO
Common Objects in Context

Keypoints Dataset

- 17 types of keypoints.
- Visibility flag annotations.
- 105,698 annotated people.
- 1,161,667 annotated keypoints.
- 45,174 unique images (train/val)

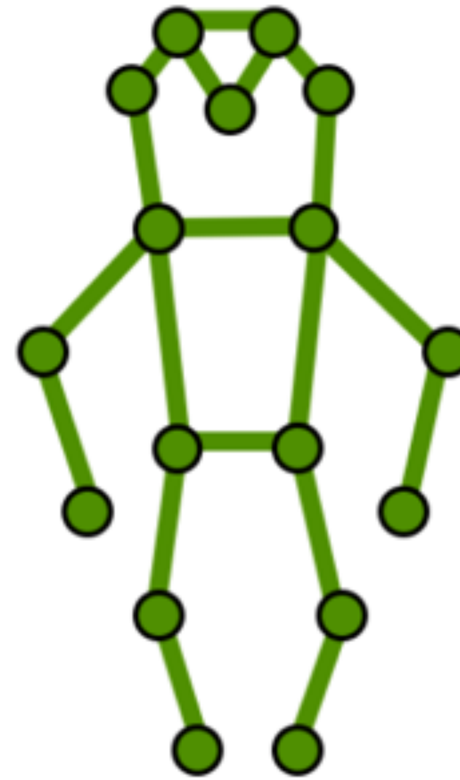


Dataset Statistics (I)

COCO
Common Objects in Context

Keypoints Dataset

- 17 types of keypoints.
- Visibility flag annotations.
- 105,698 annotated people.
- 1,161,667 annotated keypoints.
- 45,174 unique images (train/val)
- Only medium and large size people annotated



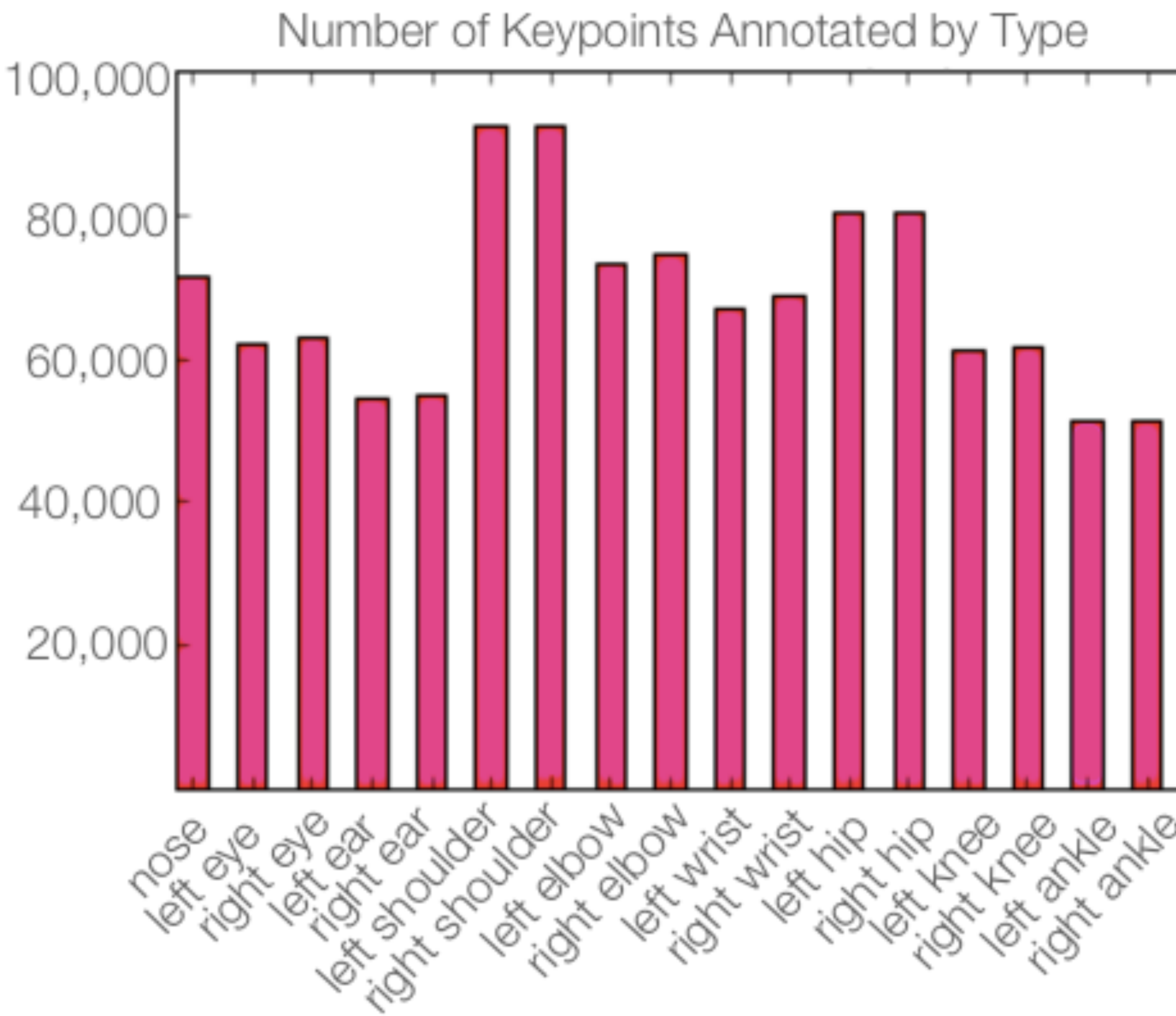
Dataset Statistics (II)

- Max people per image: **13**
- Mean people per image: **2.65**

Visibility Flag v_i

- Mean annotations per person: **11**
- 6,000 people with > 17 annotations
- 2,300 people with only 1 keypoint
- shoulders, hips, elbows, noses most popular

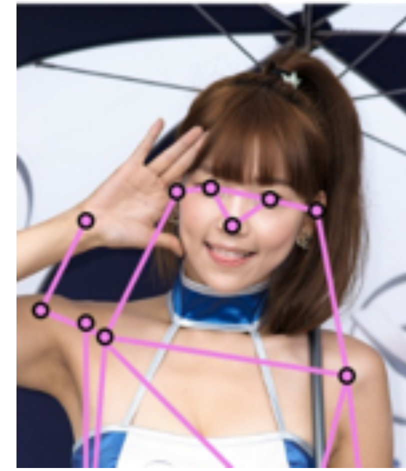
Dataset Statistics (III)

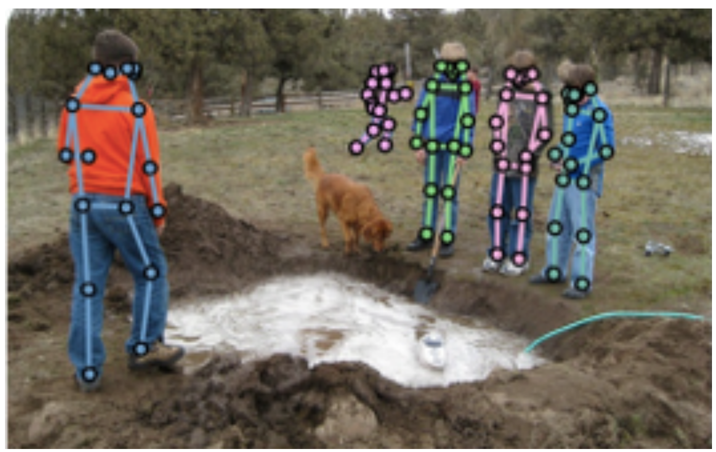


2016 COCO Keypoints Challenge

- New Task - Keypoint Detection and Localization

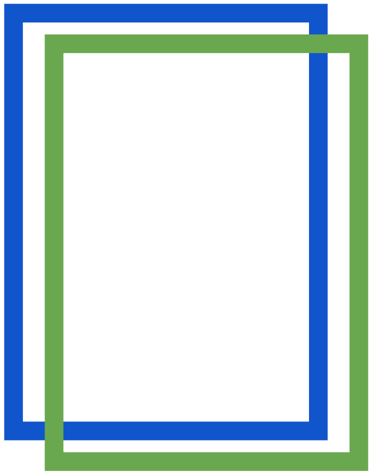
- New Metric - Object Keypoint Similarity (OKS)



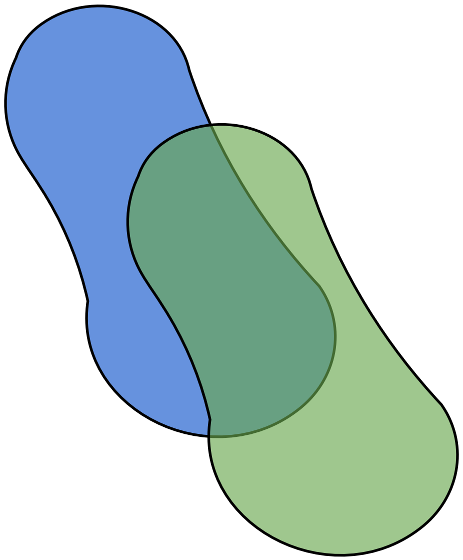


Evaluating BBoxes, Segmentations and Keypoints

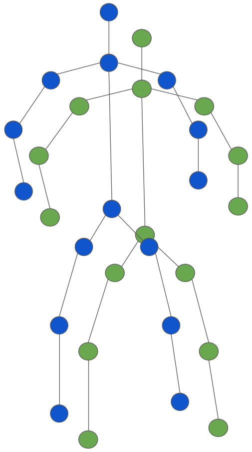
To calculate AP we need:



Bounding Box IoU

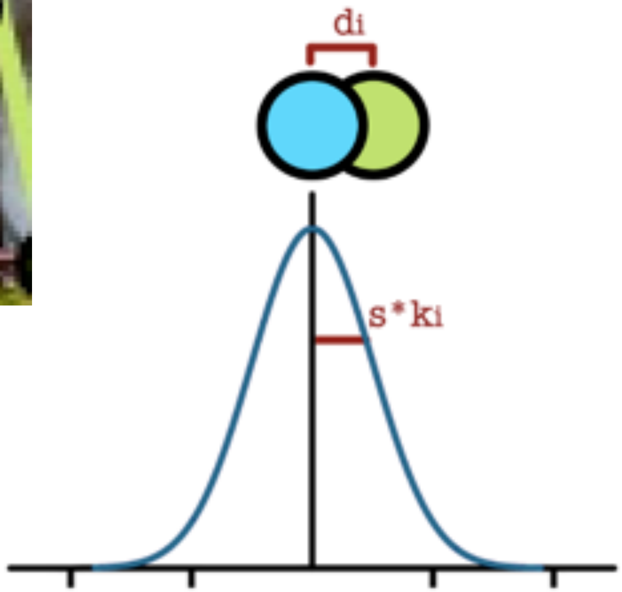


Mask IoU



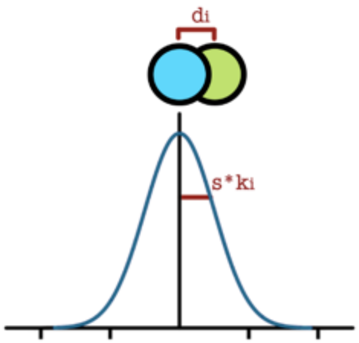
Object Keypoint Similarity

Object Keypoint Similarity -OKS



Object Keypoint Similarity

$$OKS = \frac{\sum_i e^{-\frac{d_i^2}{2s^2k_i^2}} \delta(v_i > 0)}{\sum_i \delta(v_i > 0)}$$



- d_i are the Euclidean distances between each ground truth and detected keypoint
- v_i are the visibility flags of the ground truth
- $s * k_i$(scale * keypoint constant) is the standard deviation of this gaussian
- Scale and Keypoint constant needed to equalize the importance of each keypoint
  - Eye location more precise than Hip location
- Perfect predictions will have OKS=1
- Predictions with keypoints wrong by more than a few standard deviations will have OKS~0

Scale and Standard Deviation

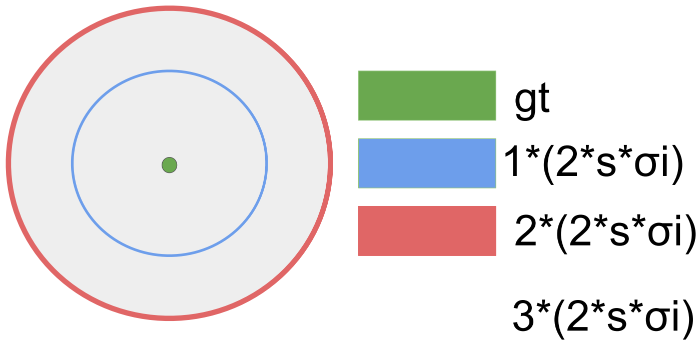
s scale

k_i keypoint

Standard Deviation per Keypoint type

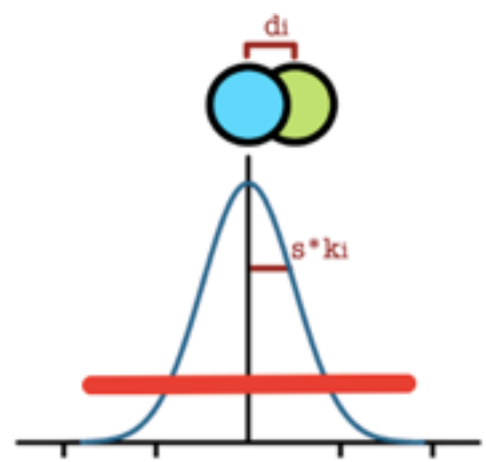
Keypoint	k_i
hips	0.107
ankles	0.089
knees	0.087
shoulders	0.079
elbows	0.072
wrists	0.062
ears	0.035
nose	0.026
eyes	0.025

OKS Cut-off



OKS = 1
OKS = .88
OKS = .61
OKS = .32

- Minimum OKS value is 0.5
- We analyze only keypoints that lie within ~2.77 of their standard deviation



From OKS to COCO AP

From OKS to COCO AP

COCO AP is the area under the PR curve

From OKS to COCO AP

COCO AP is the area under the PR curve

precision

1
0.8
0.6
0.4
0.2

precision

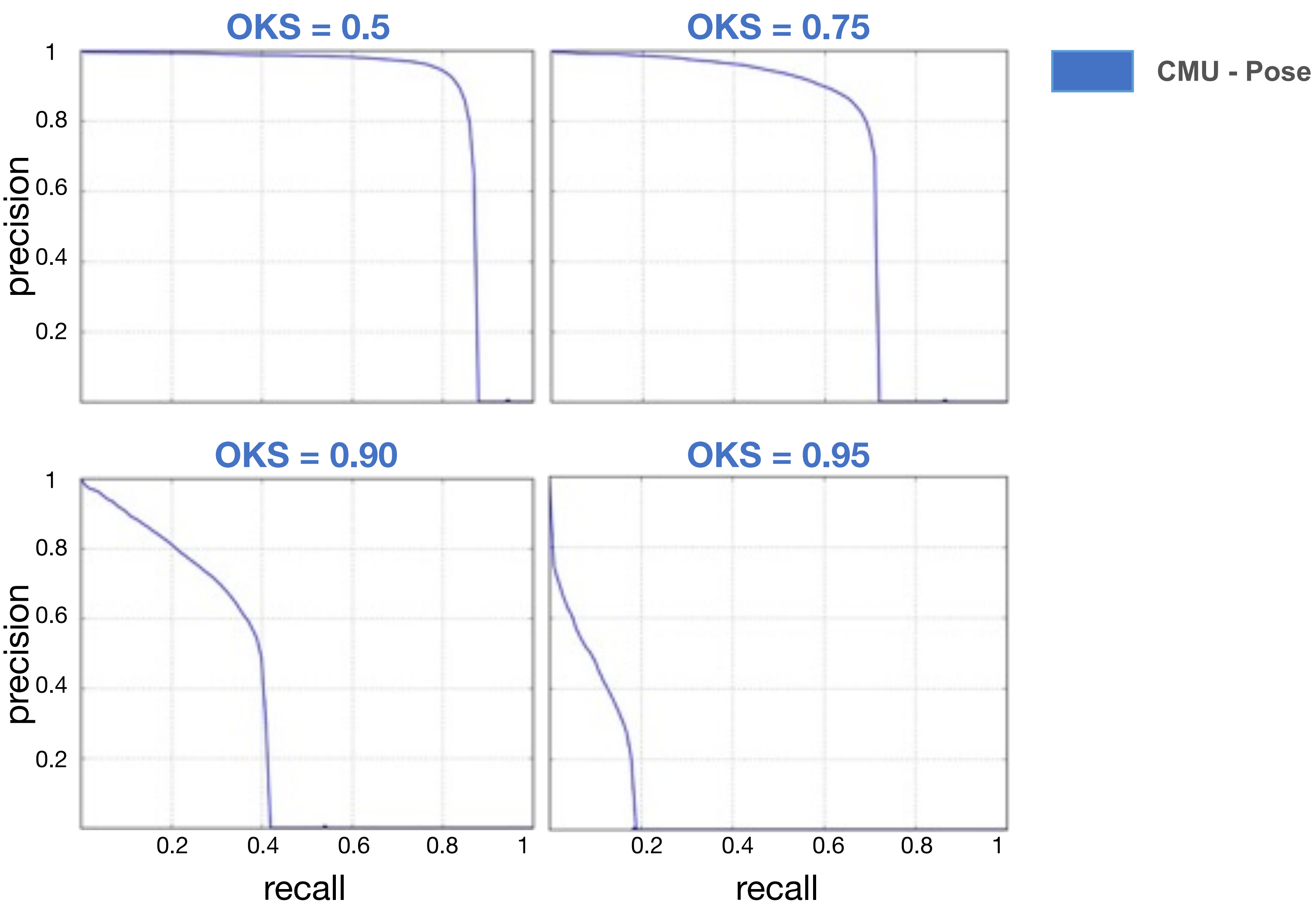
1
0.8
0.6
0.4
0.2

0.2 0.4 0.6 0.8 1
recall

0.2 0.4 0.6 0.8 1
recall

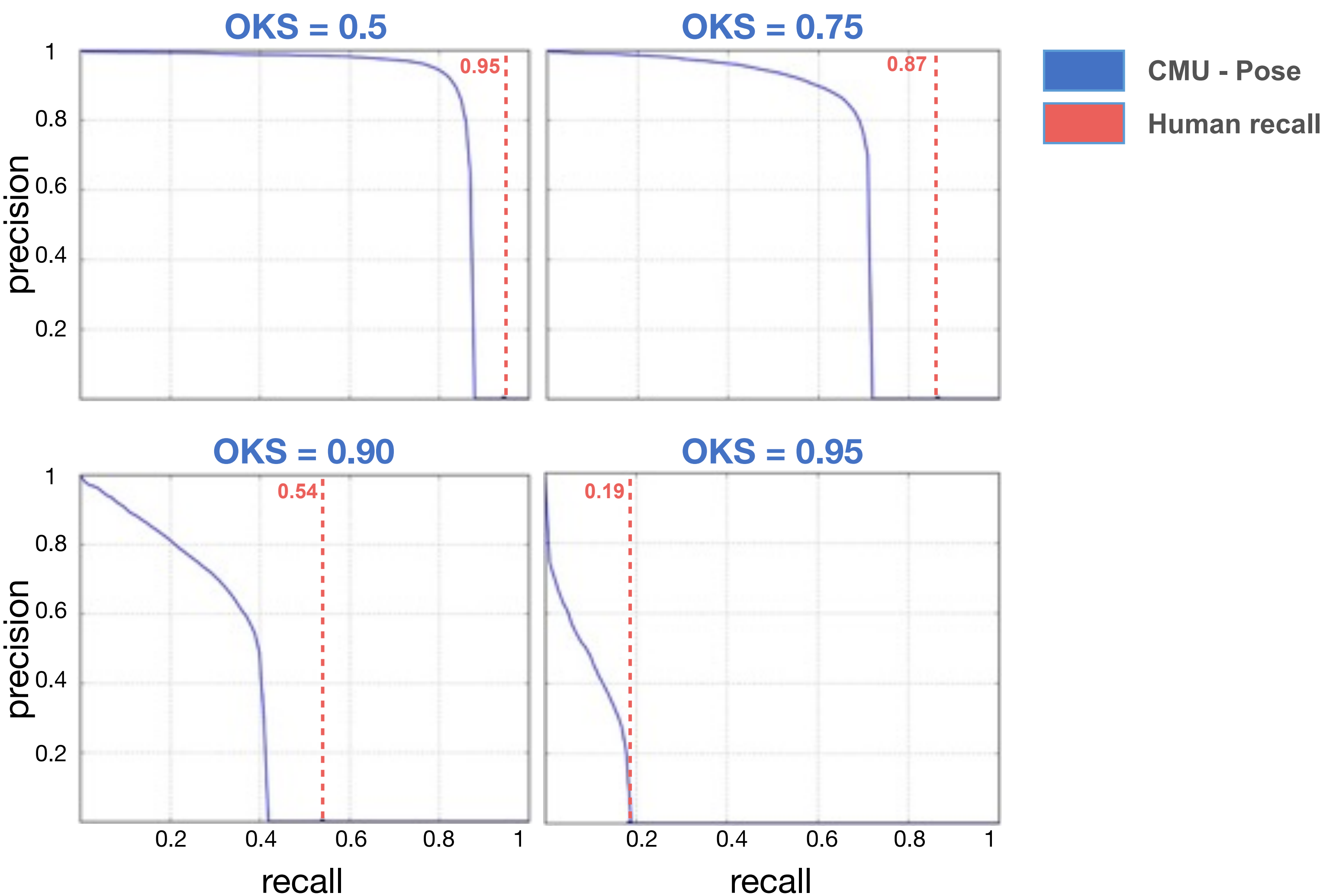
From OKS to COCO AP

COCO AP is the area under the PR curve



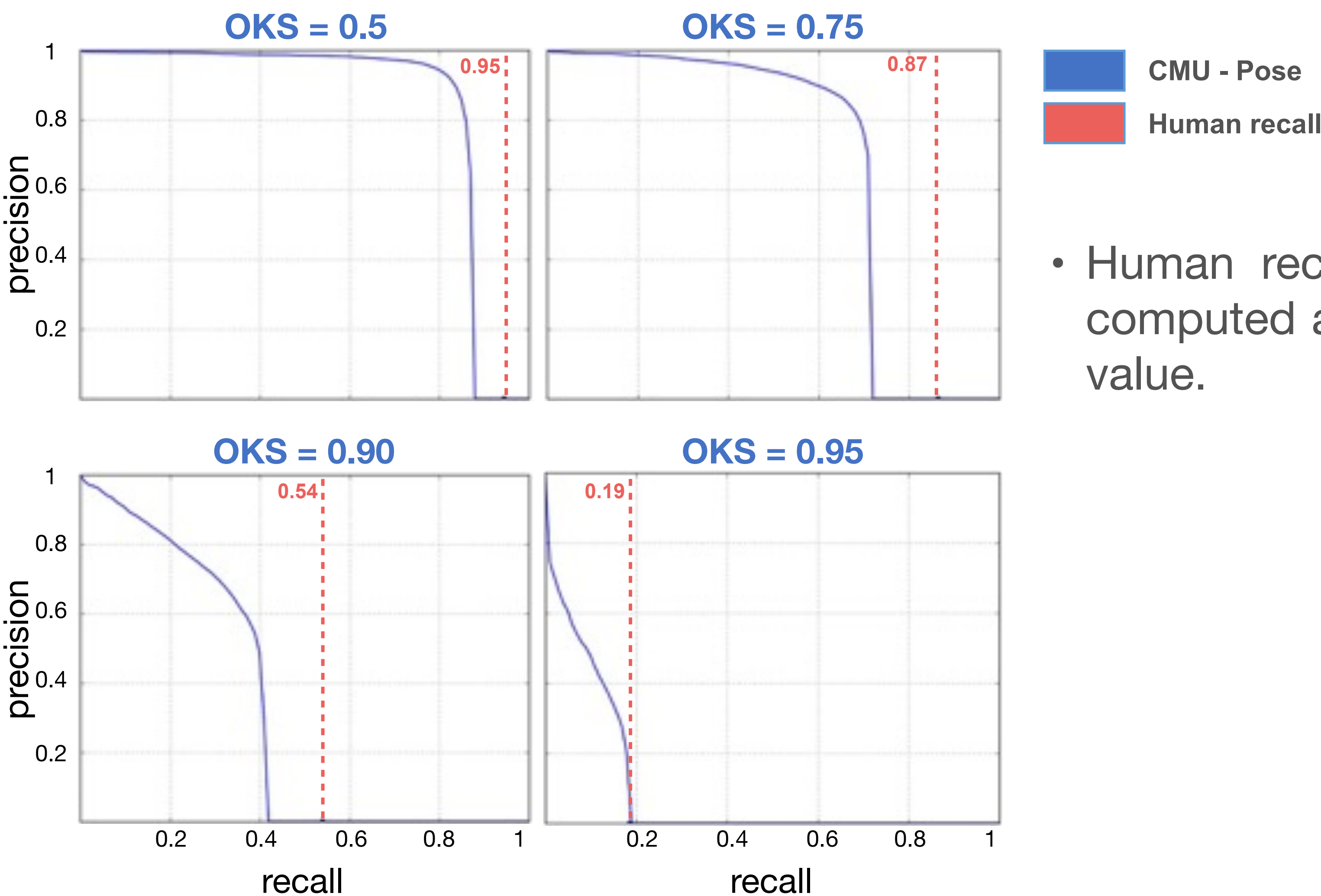
From OKS to COCO AP

COCO AP is the area under the PR curve



From OKS to COCO AP

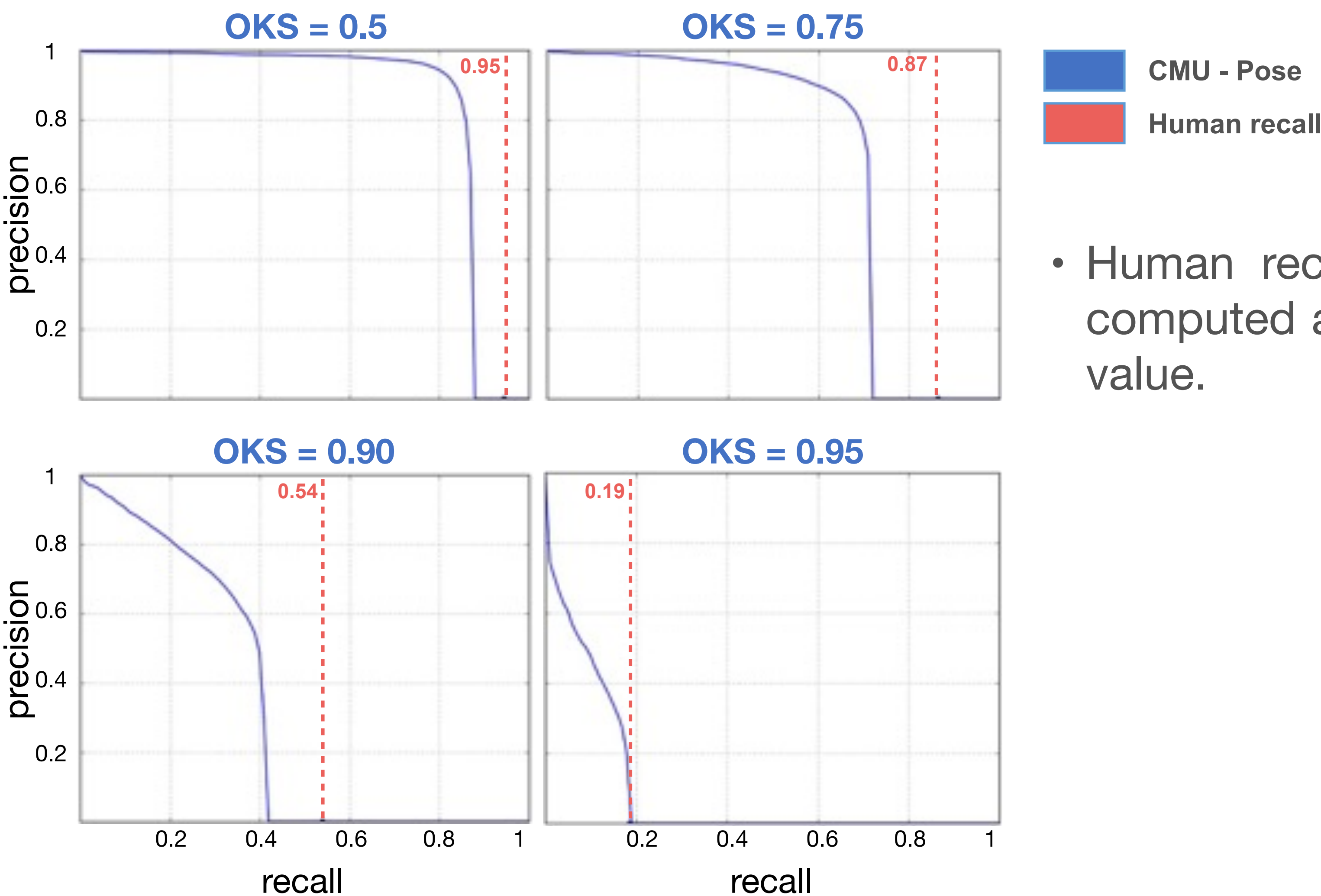
COCO AP is the area under the PR curve



- Human recall can be computed at any OKS value.

From OKS to COCO AP

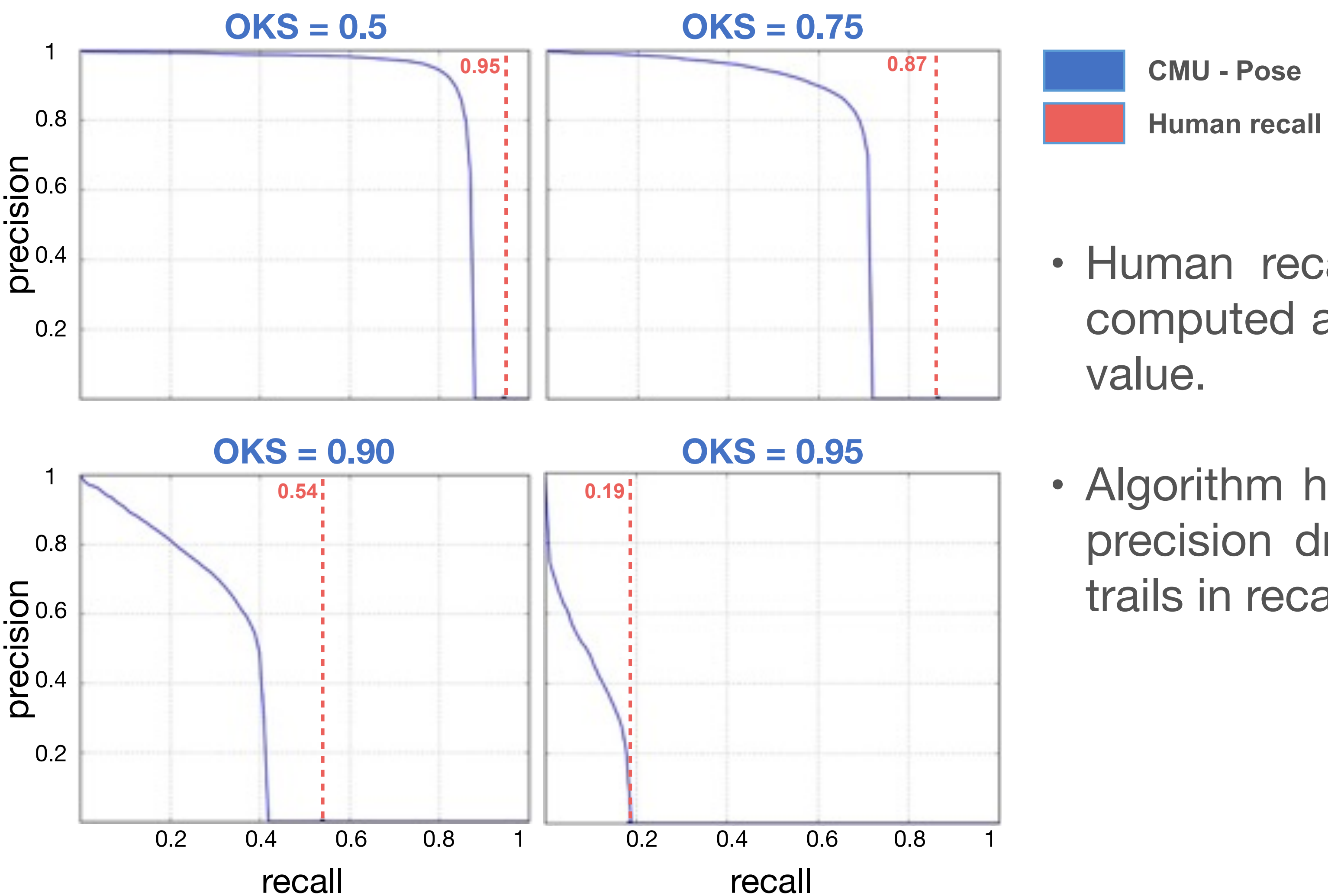
COCO AP is the area under the PR curve



- Human recall can be computed at any OKS value.

From OKS to COCO AP

COCO AP is the area under the PR curve



- Human recall can be computed at any OKS value.
- Algorithm has smooth precision drop-off but trails in recall.

COCO Challenge Results

COCO
Common Objects in Context

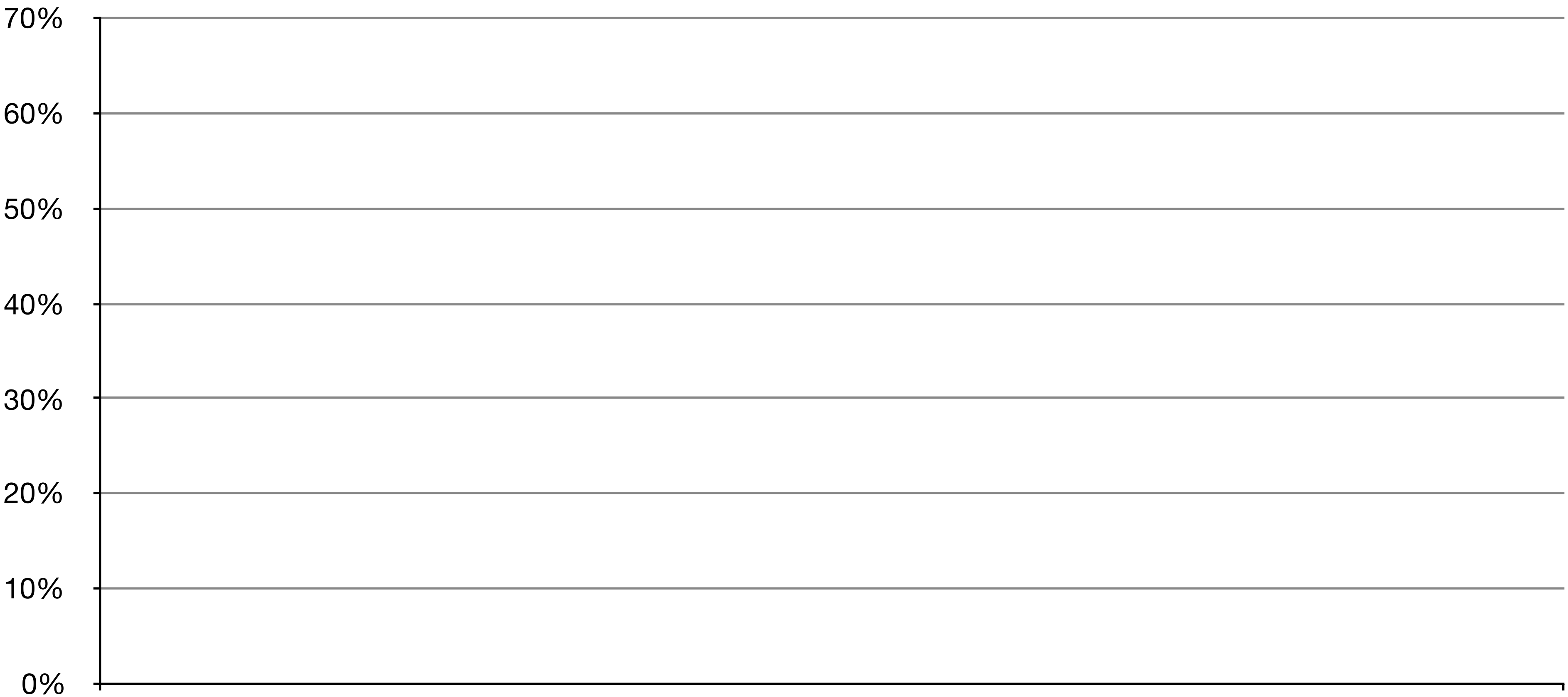
Keypoints

Keypoints Leaderboard (I)

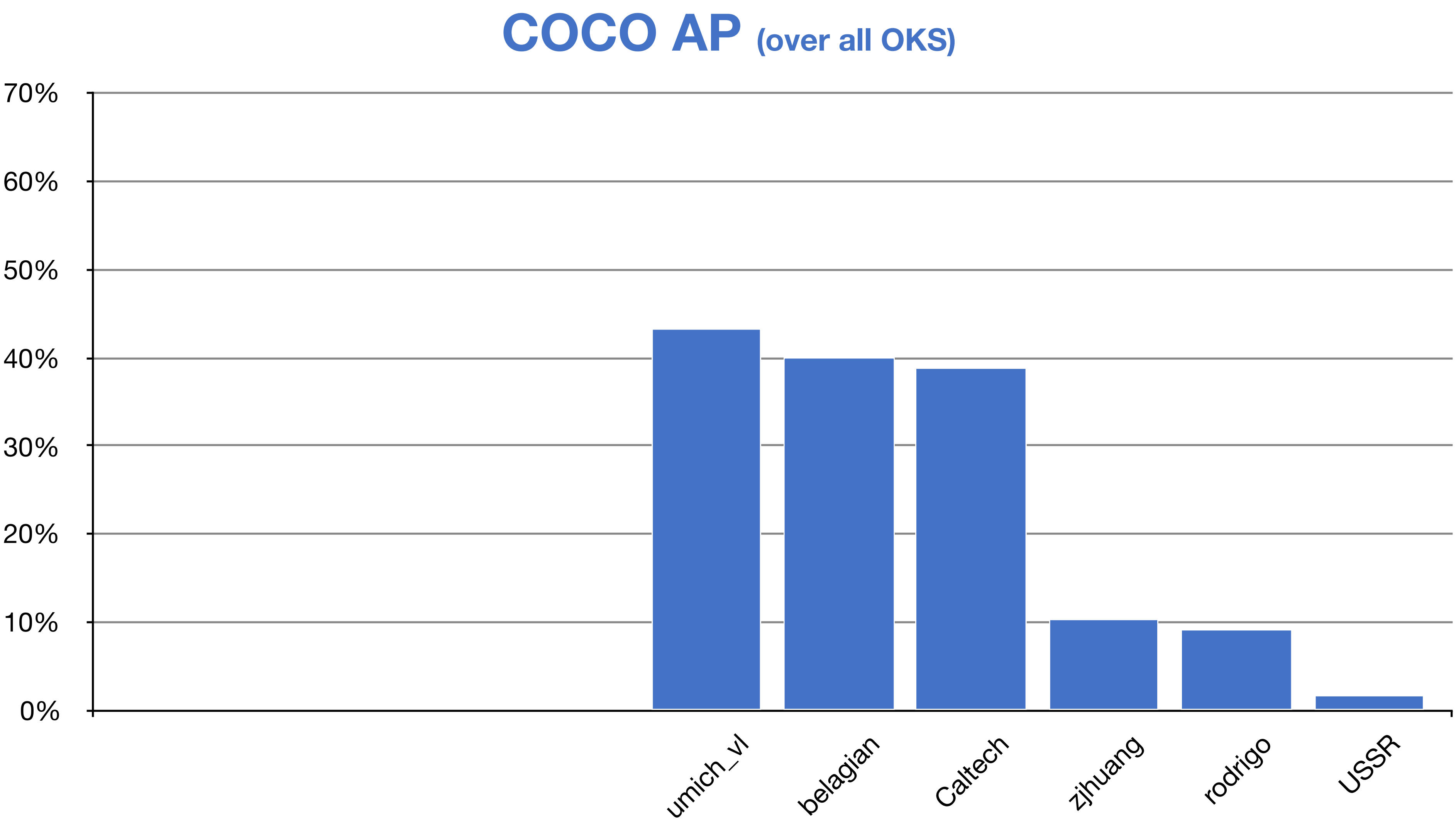
COCO AP (over all OKS)

Keypoints Leaderboard (I)

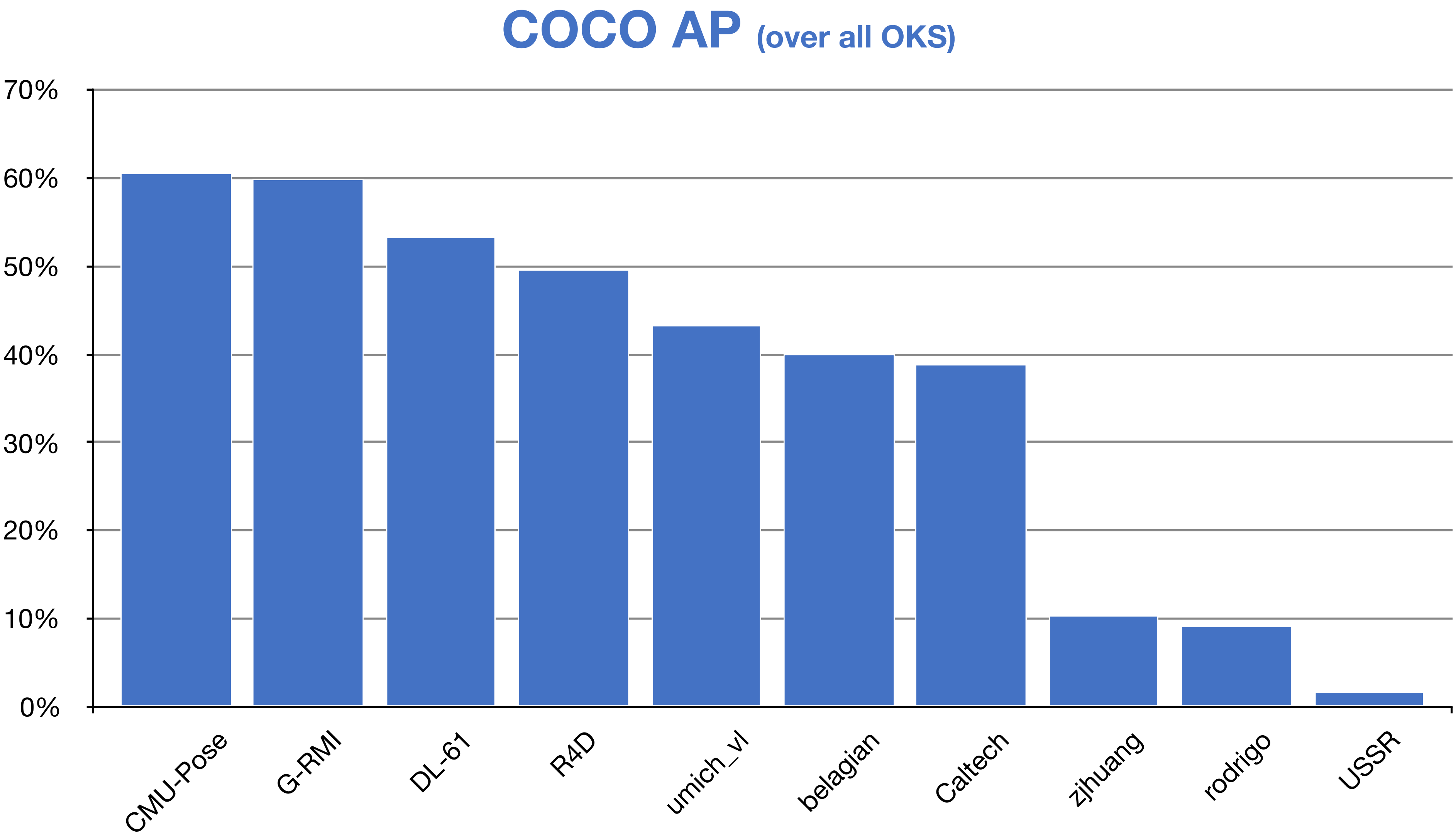
COCO AP (over all OKS)



Keypoints Leaderboard (I)



Keypoints Leaderboard (I)

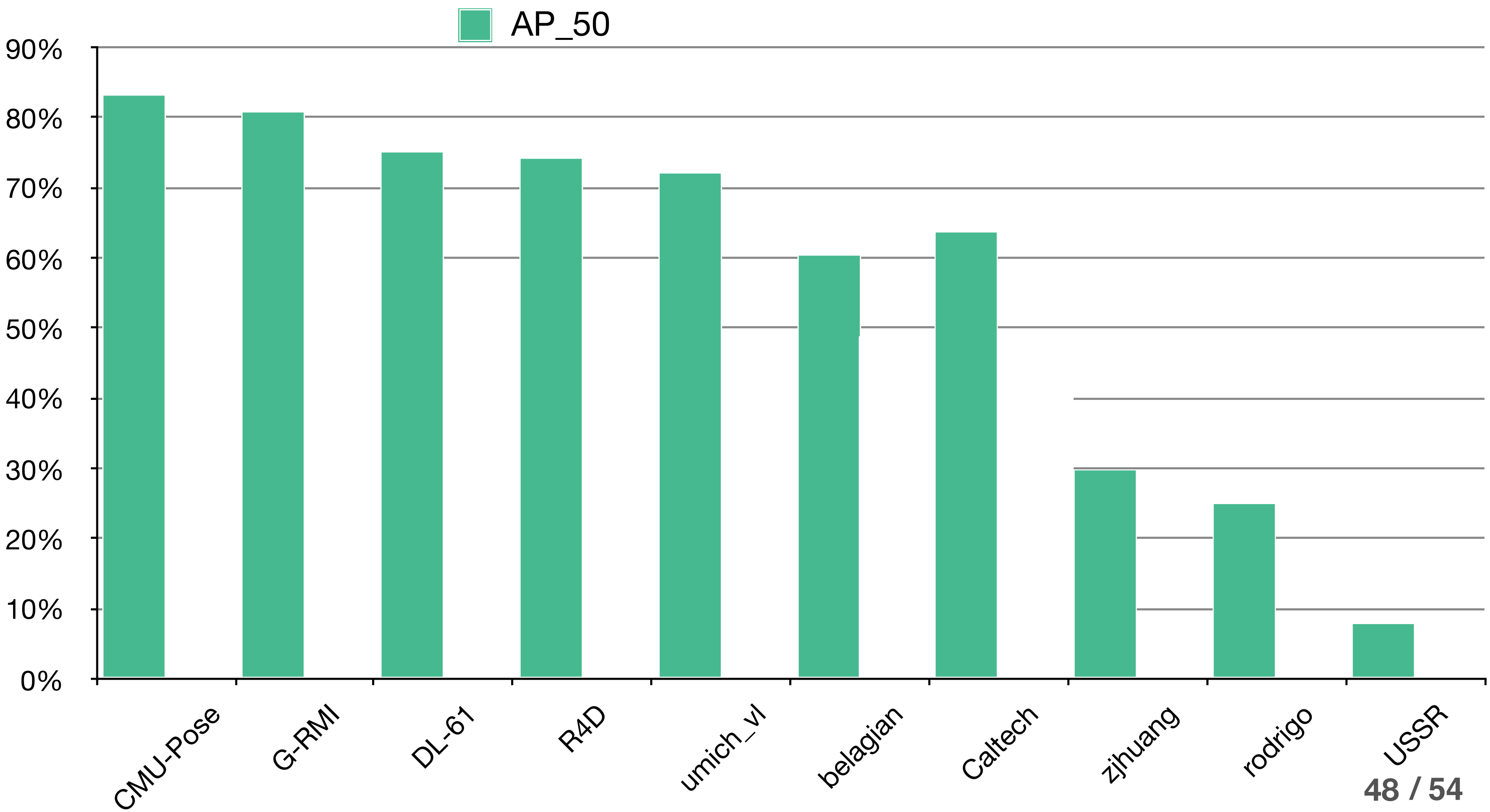


Keypoints Leaderboard (II)

Instance localization affects also keypoints

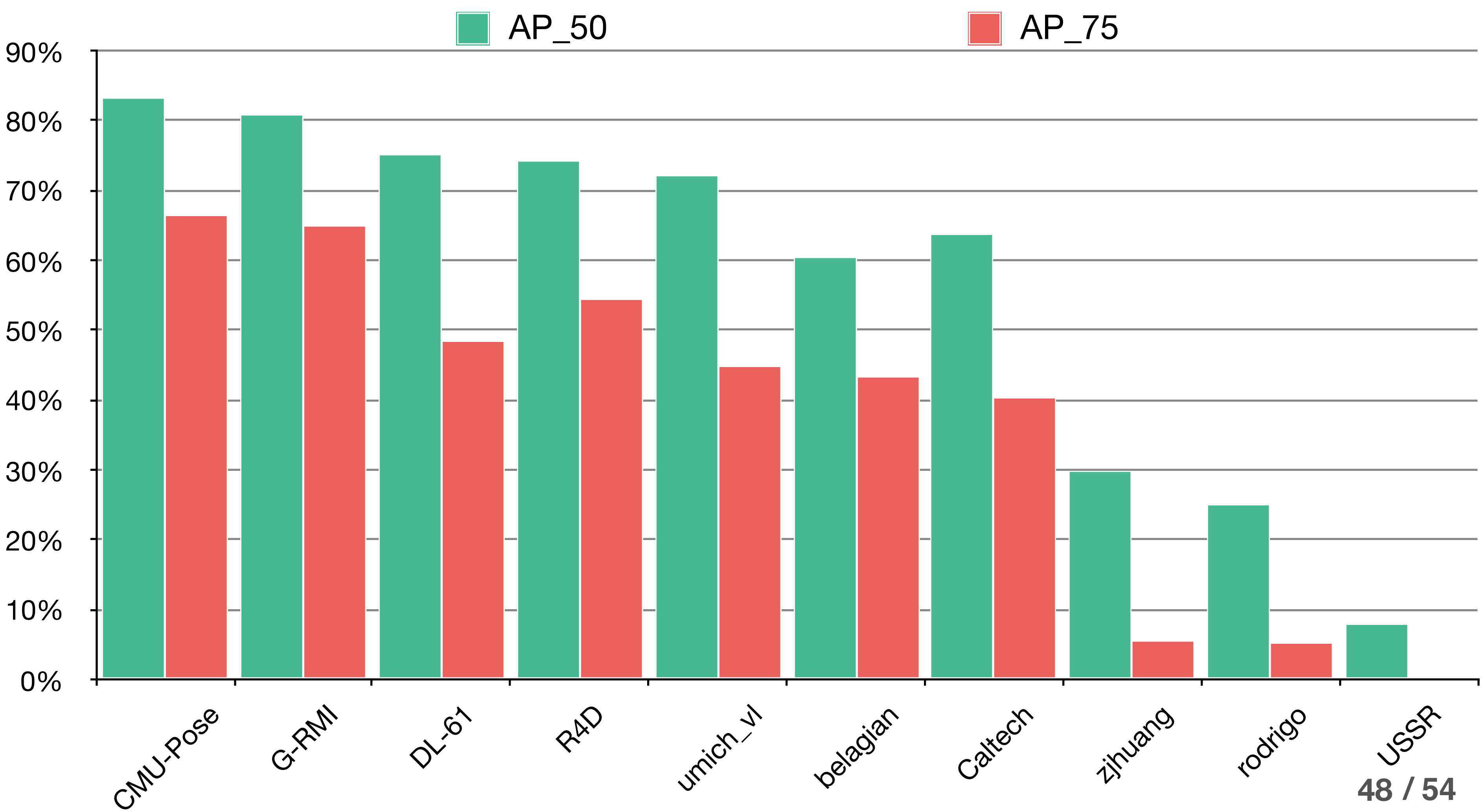
Keypoints Leaderboard (II)

Instance localization affects also keypoints



Keypoints Leaderboard (II)

Instance localization affects also keypoints

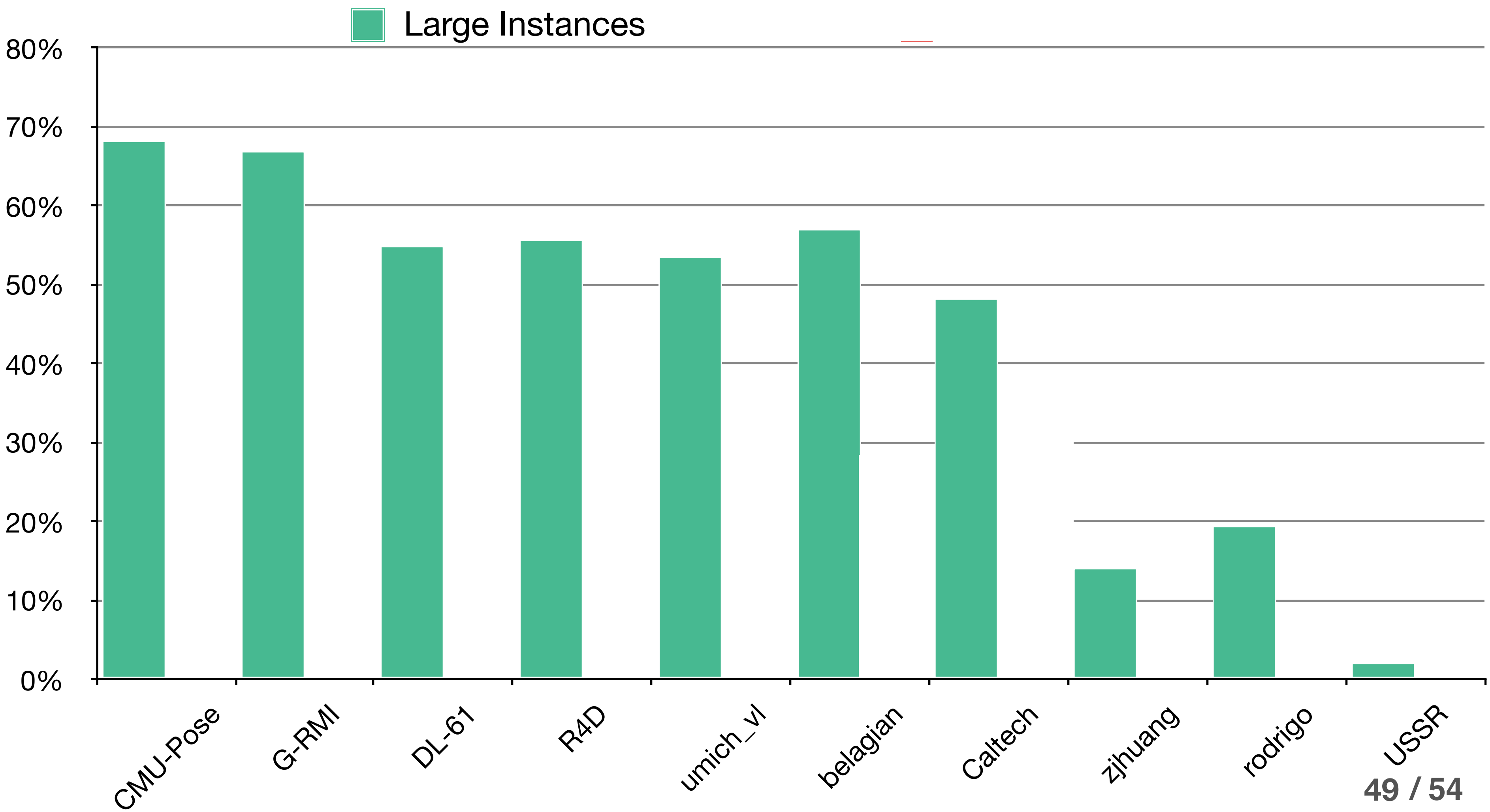


Keypoints Leaderboard (III)

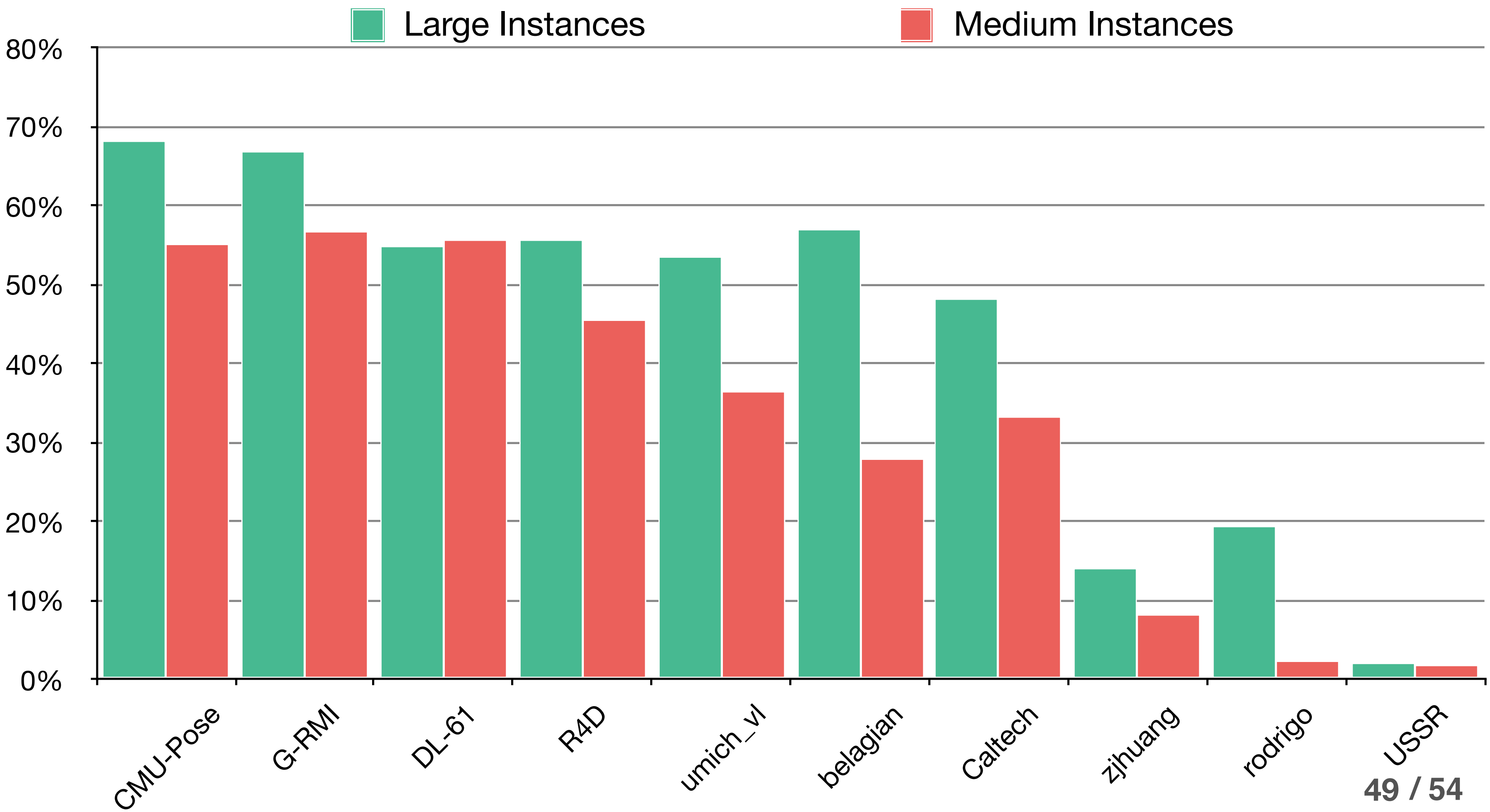
Instance scale is an important factor

Keypoints Leaderboard (III)

Instance scale is an important factor



Instance scale is an important factor



Performance Breakdown (I)

COCO AP varies across keypoints

CMU-Pose G-RMI DL-61 R4D
umich_vl belagian Caltech

100%

75%

50%

25%

0%

Performance Breakdown (I)

COCO AP varies across keypoints

- CMU-Pose
- G-RMI
- DL-61
- R4D
- umich_vl
- belagian
- Caltech

100%

75%

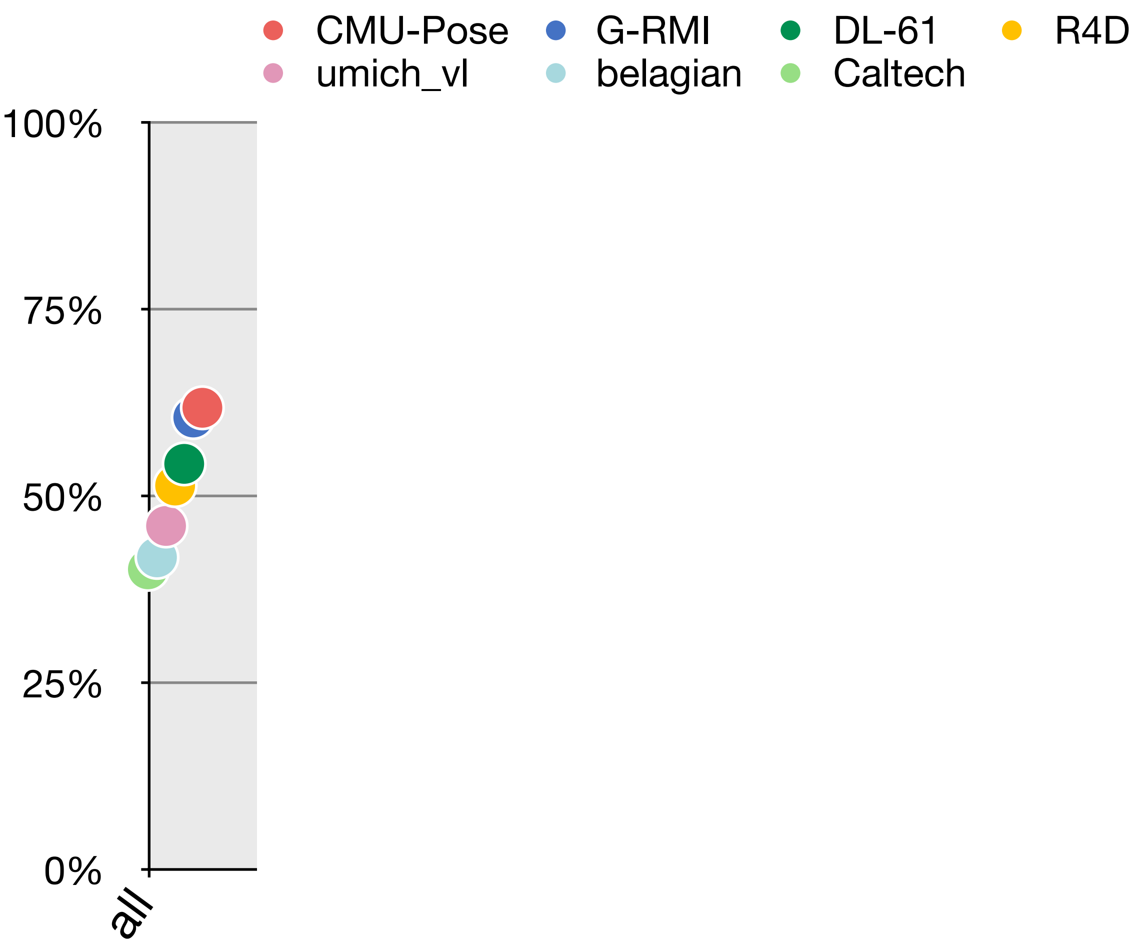
50%

25%

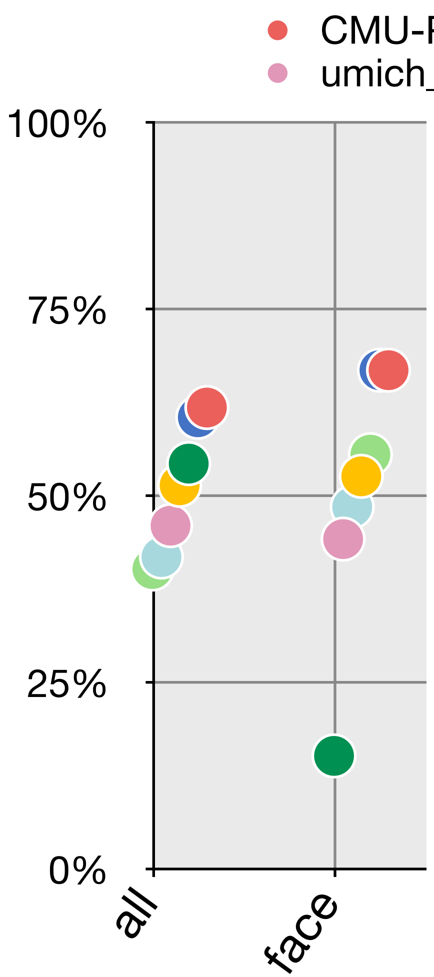
0%

Performance Breakdown (I)

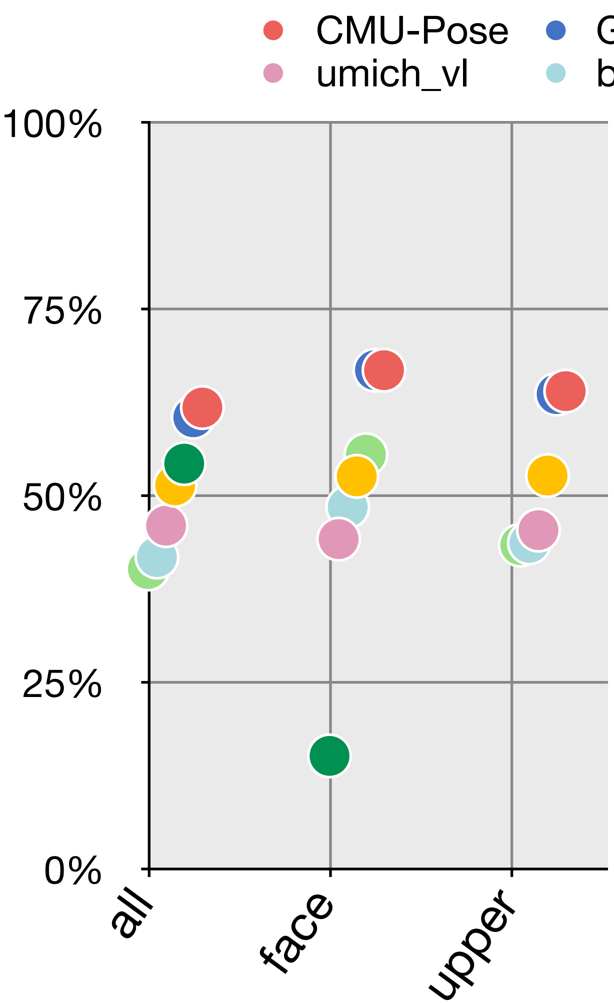
COCO AP varies across keypoints

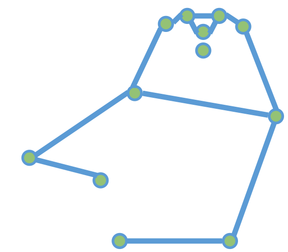


COCO AP varies across keypoints

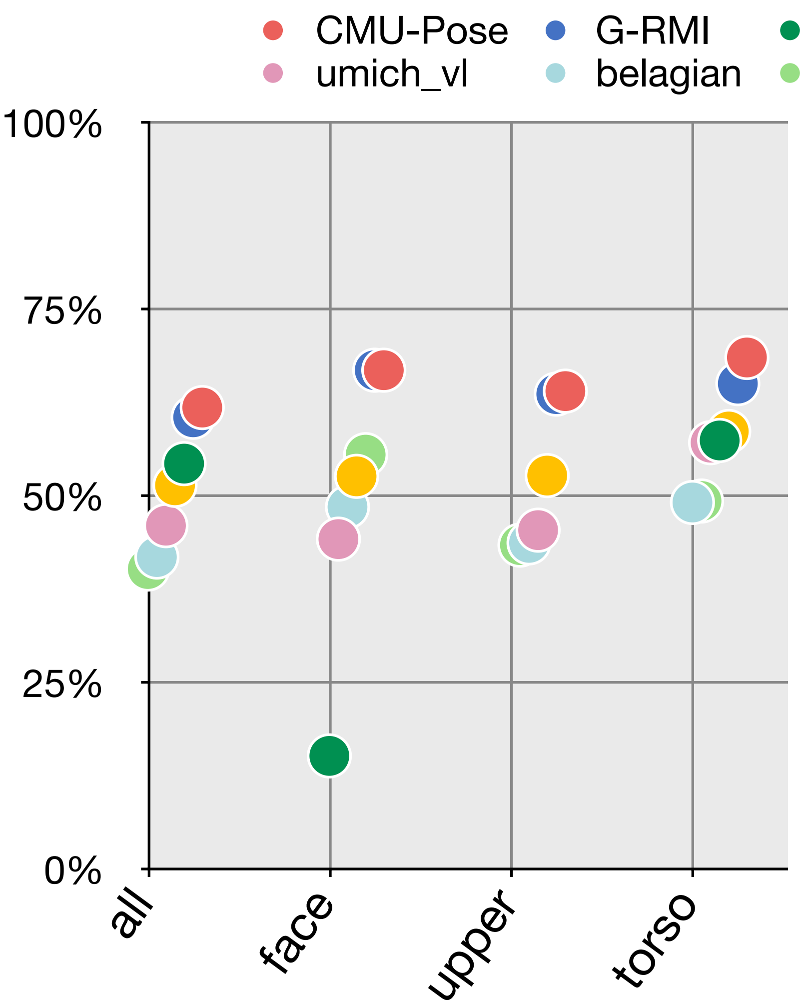


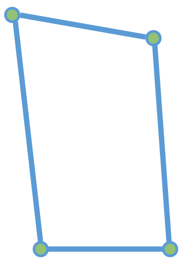
COCO AP varies across keypoints





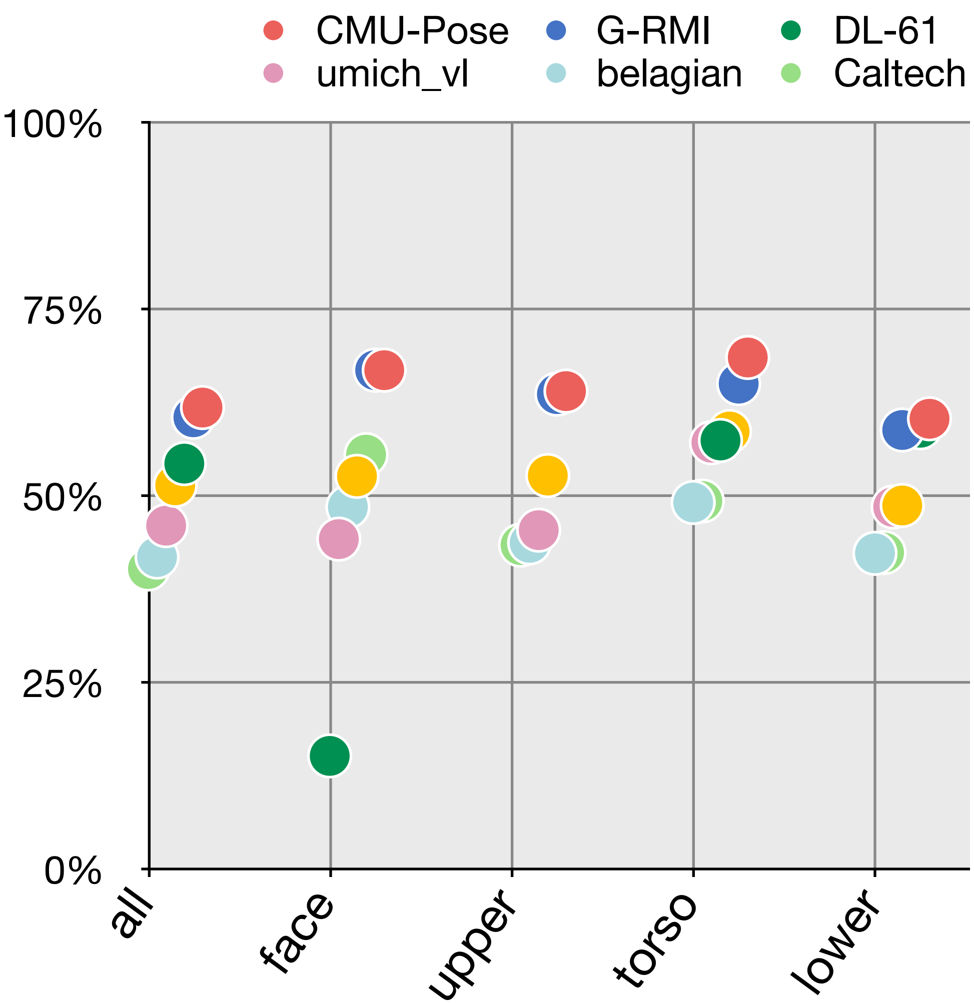
COCO AP varies across keypoints

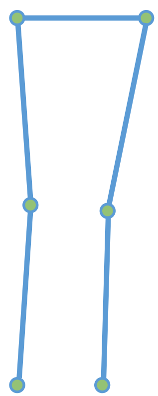




Performance Breakdown (I)

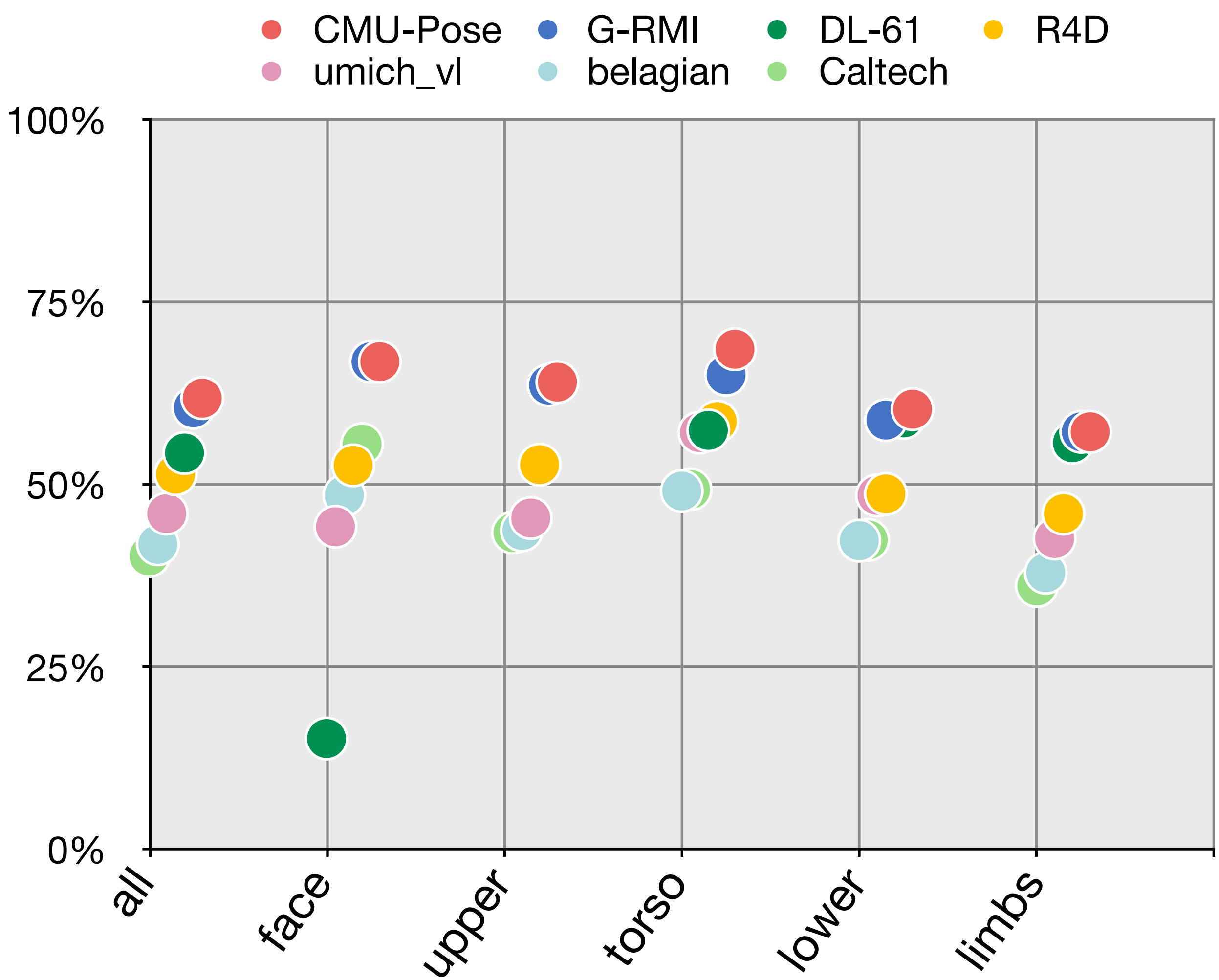
COCO AP varies across keypoints

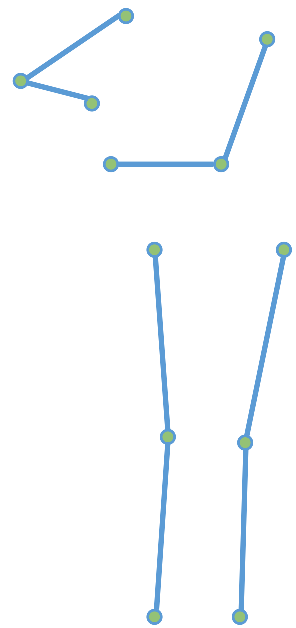




Performance Breakdown (I)

COCO AP varies across keypoints





Correlation between methods

Correlation between methods

How similarly do algorithms perform?

Correlation between methods

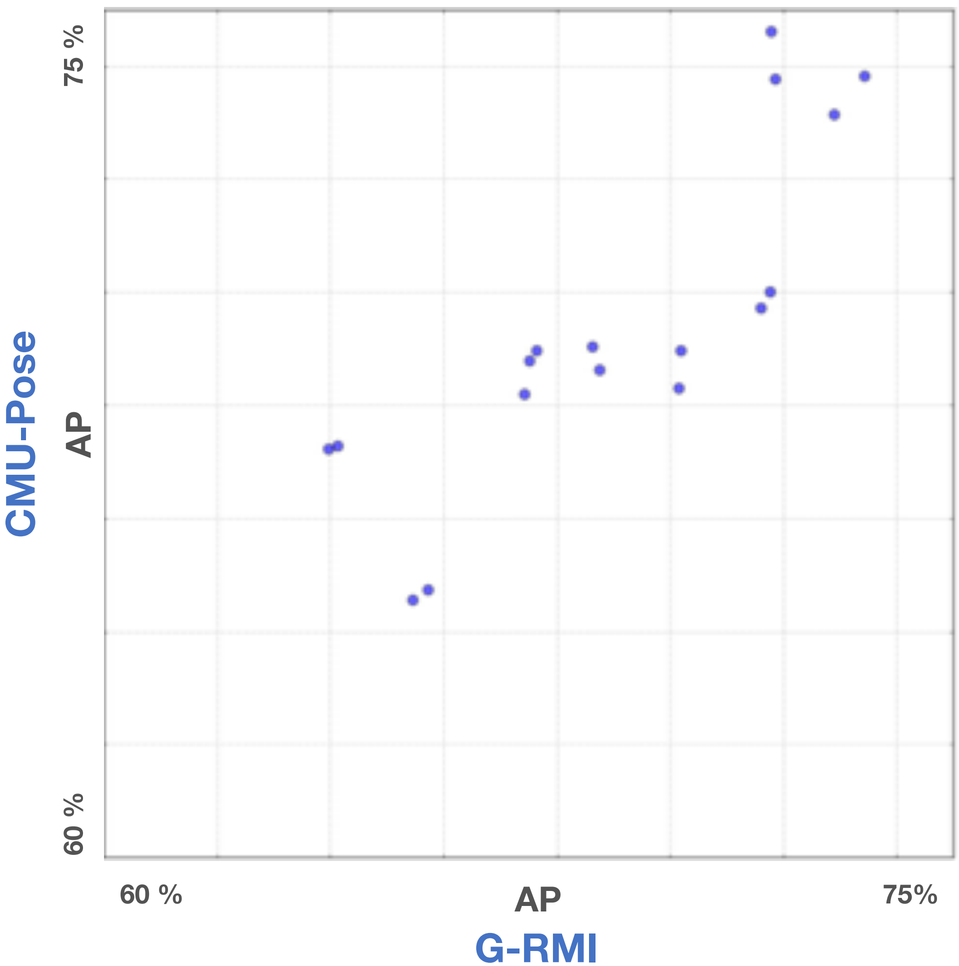
How similarly do algorithms perform?

CMU-Pose

G-RMI

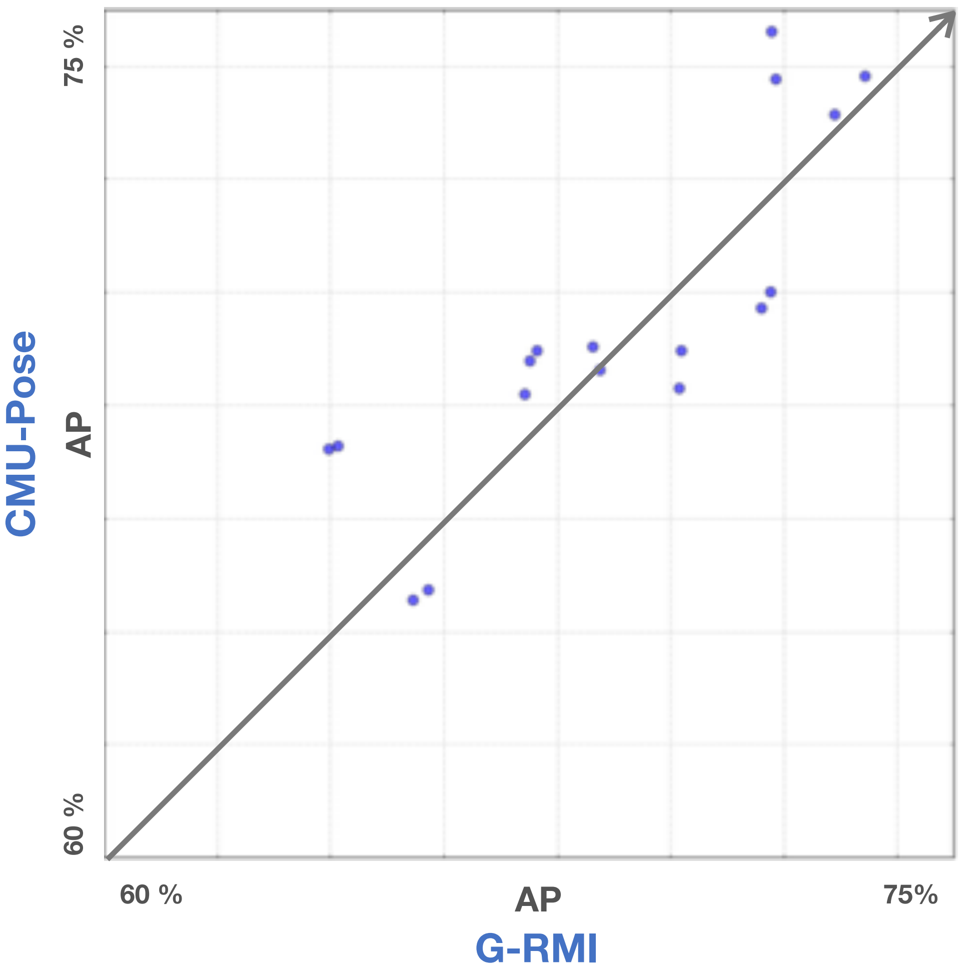
Correlation between methods

How similarly do algorithms perform?



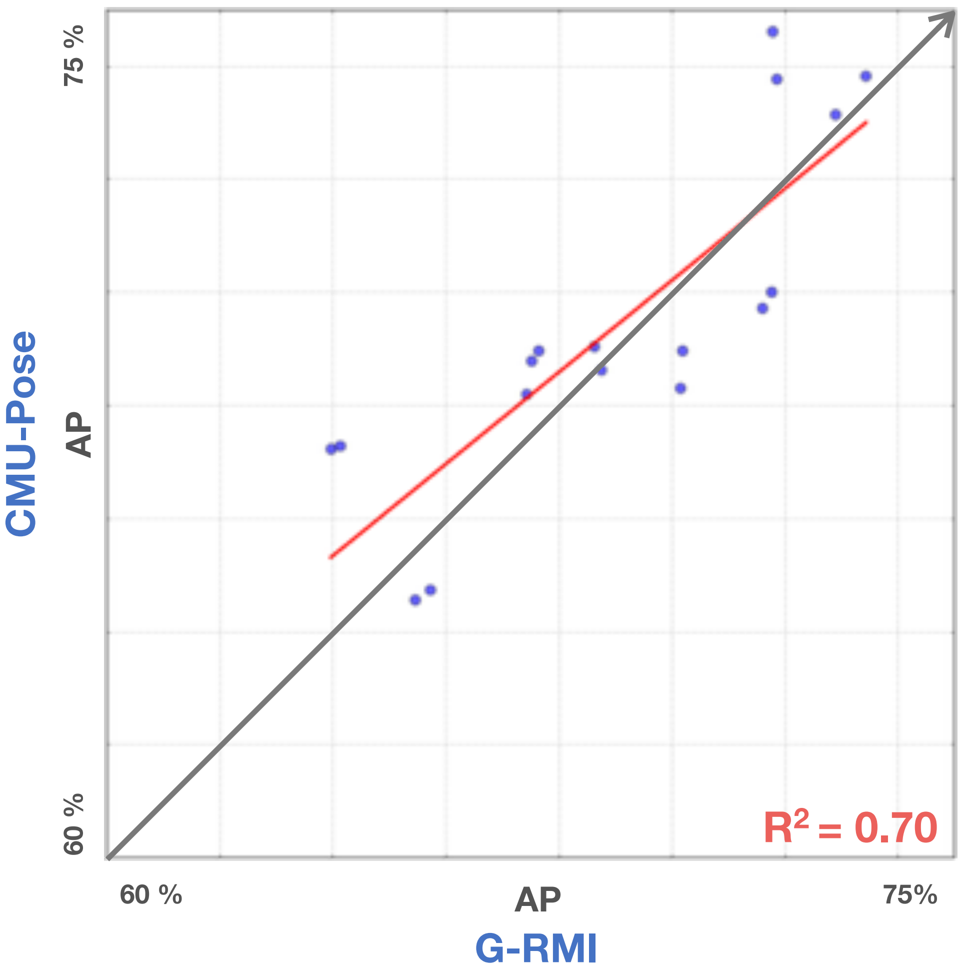
Correlation between methods

How similarly do algorithms perform?



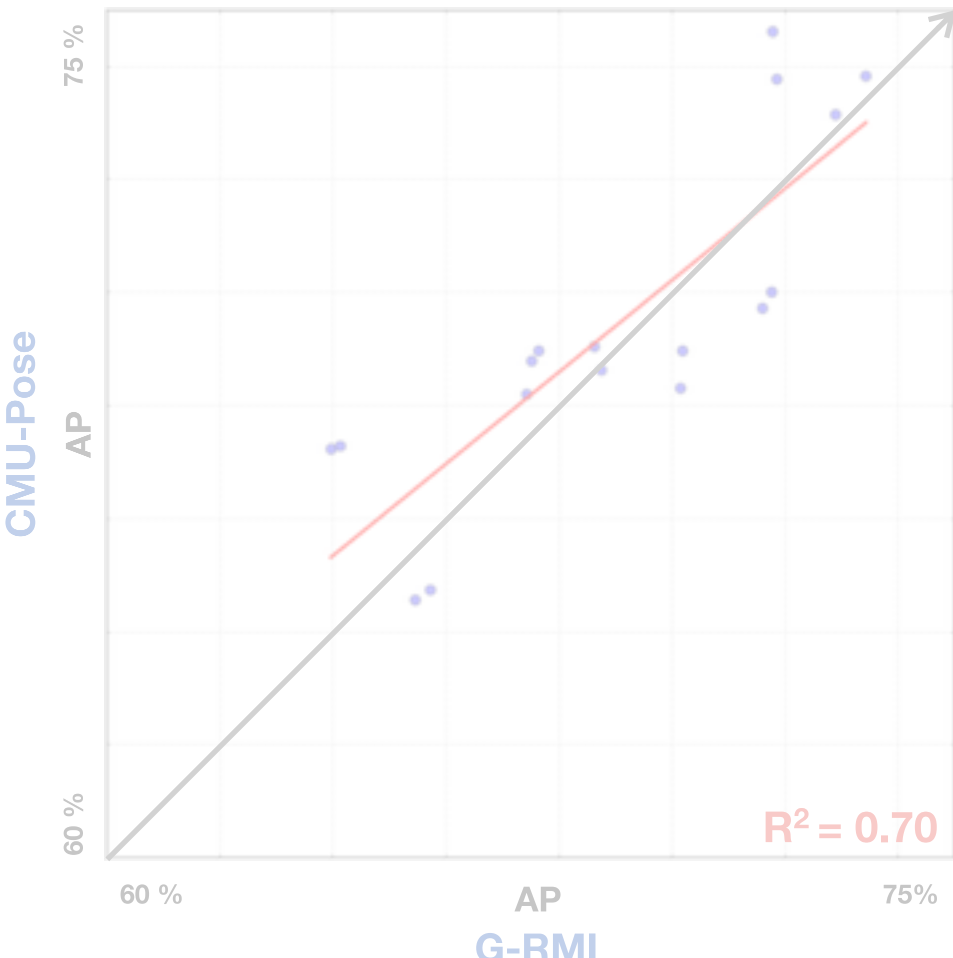
Correlation between methods

How similarly do algorithms perform?



Correlation between methods

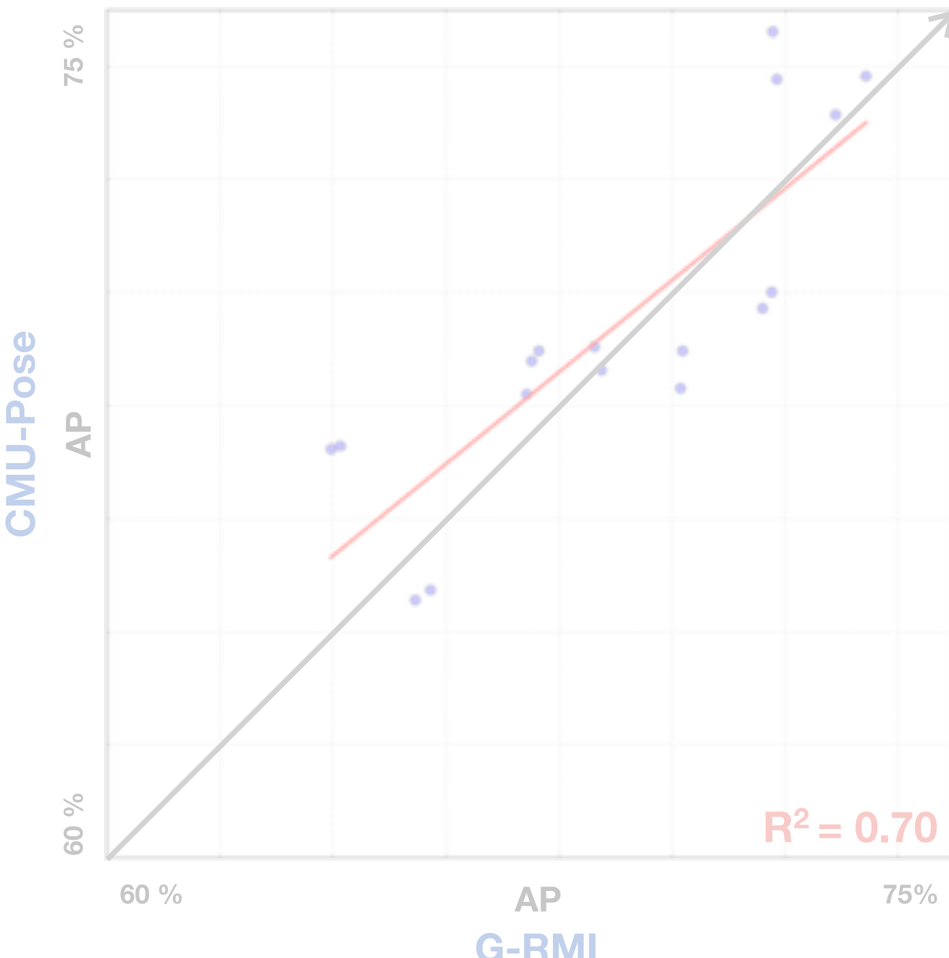
How similarly do algorithms perform?

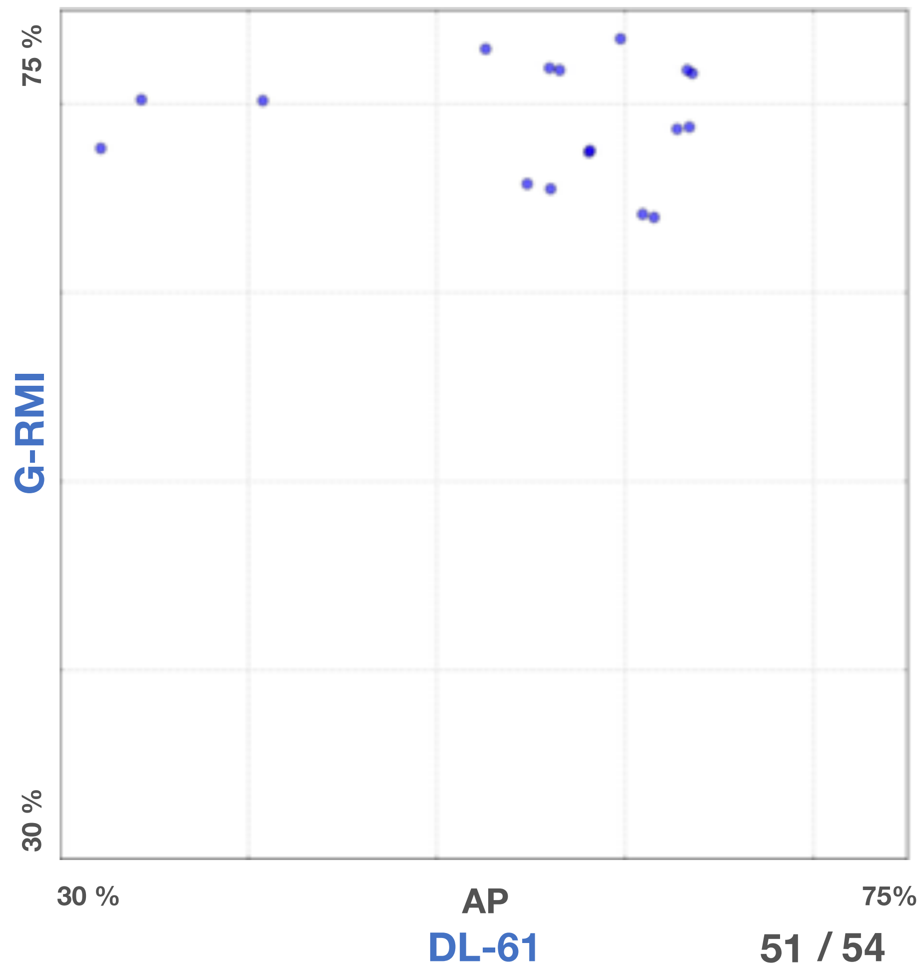


G-RMI

Correlation between methods

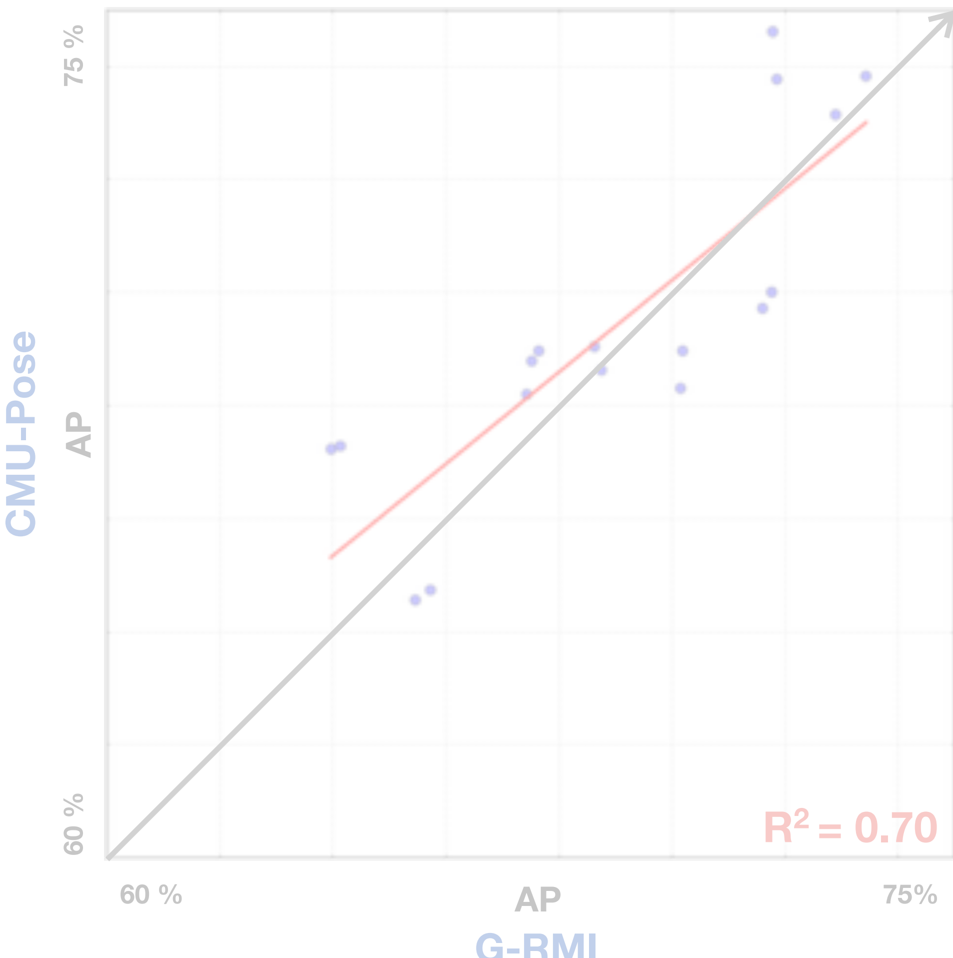
How similarly do algorithms perform?

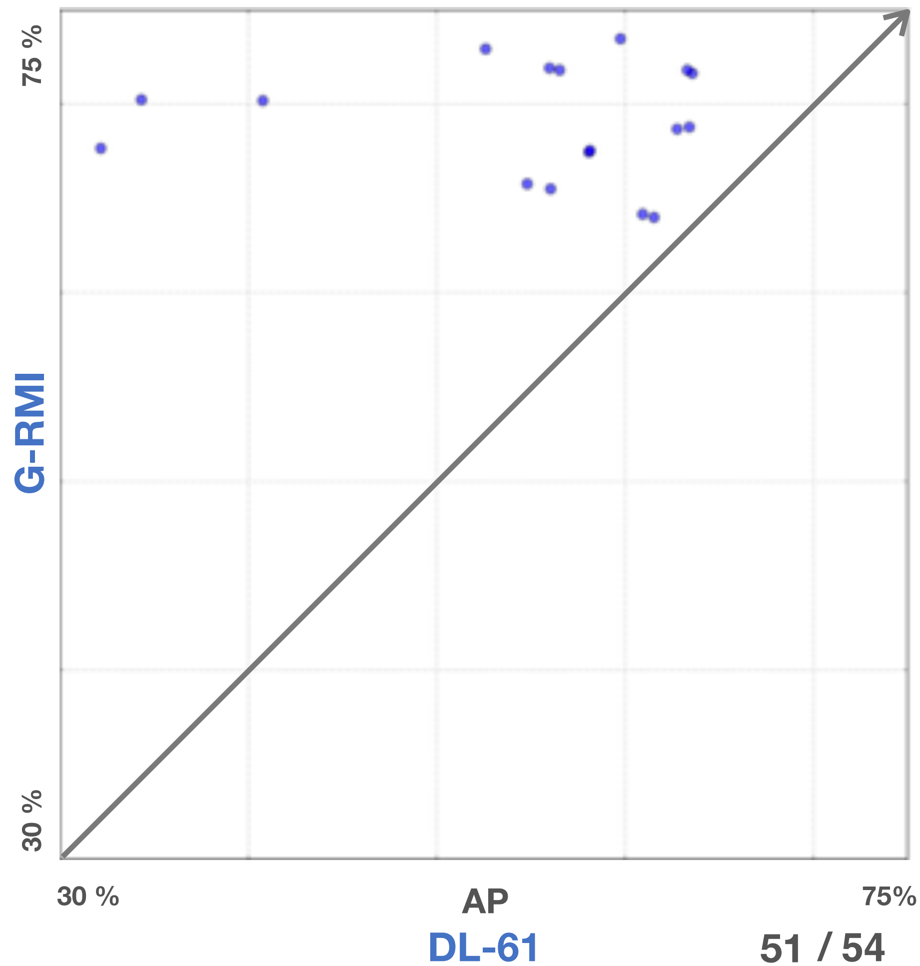




Correlation between methods

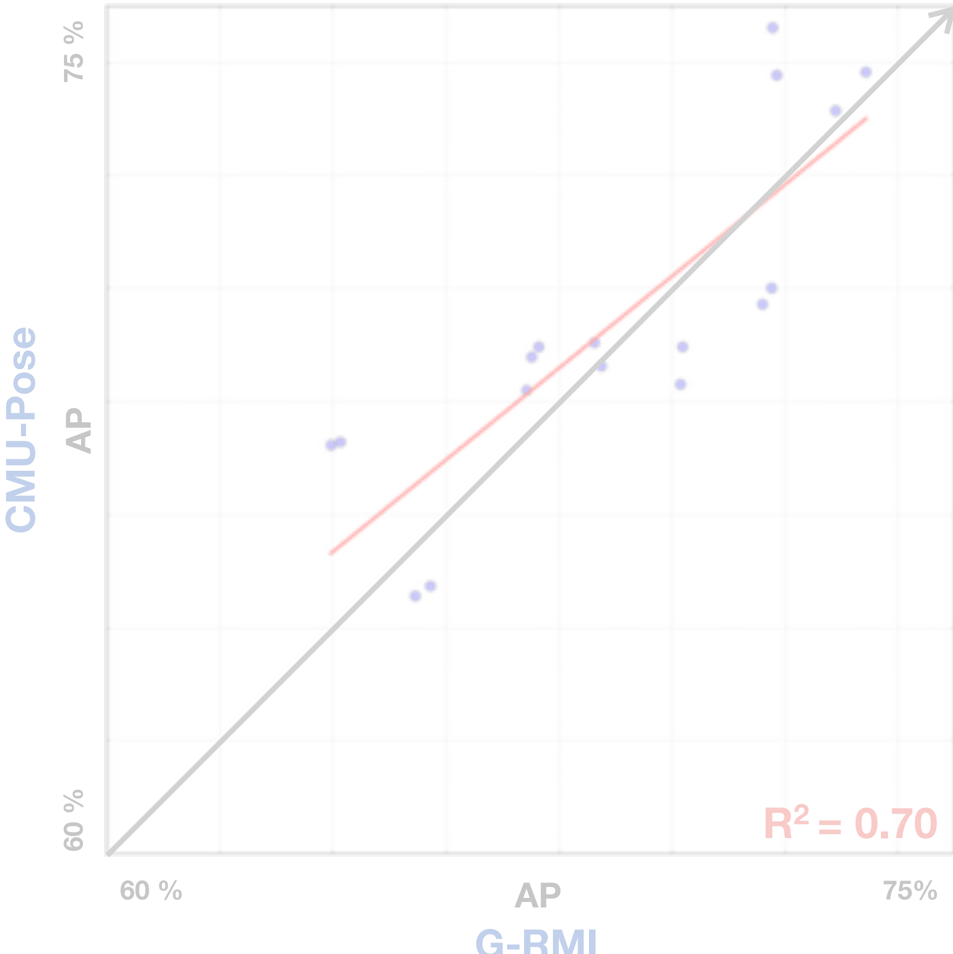
How similarly do algorithms perform?

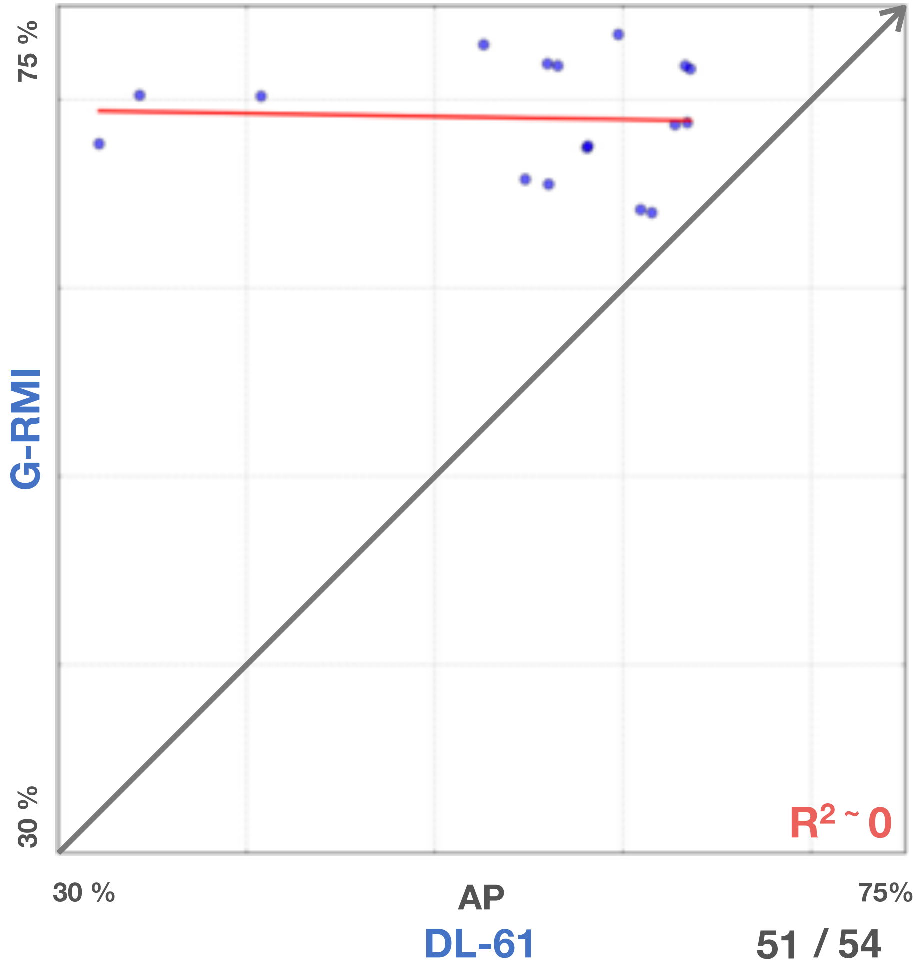




Correlation between methods

How similarly do algorithms perform?





Types of Keypoint Errors

Types of Keypoint Errors

Excusable false positives.

Types of Keypoint Errors

Errors with high OKS impact.

Types of Keypoint Errors

Errors with low OKS impact.

Summary of Findings

Summary of Findings

2016 Keypoint Challenge Take-aways

2016 Keypoint Challenge Take-aways

- OKS is a new distance metric designed to equalize contribution of all keypoints.

Summary of Findings

2016 Keypoint Challenge Take-aways

- OKS is a new distance metric designed to equalize contribution of all keypoints.
- Localization errors are predominant.

Summary of Findings

2016 Keypoint Challenge Take-aways

- OKS is a new distance metric designed to equalize contribution of all keypoints.
- Localization errors are predominant.
- Size of instances is impactful on performance.

Summary of Findings

2016 Keypoint Challenge Take-aways

- OKS is a new distance metric designed to equalize contribution of all keypoints.
- Localization errors are predominant.
- Size of instances is impactful on performance.
- Performance can vary greatly across different keypoint sets.

Summary of Findings

2016 Keypoint Challenge Take-aways

- OKS is a new distance metric designed to equalize contribution of all keypoints.
- Localization errors are predominant.
- Size of instances is impactful on performance.
- Performance can vary greatly across different keypoint sets.
- Human baselines are coming soon!

Challenge Ranking

Challenge Ranking

Team	Keypoints
CMU-Pose	1st
G-RMI	2nd
DL-61	3rd

Challenge Ranking

Team	Keypoints
CMU-Pose	1st
G-RMI	2nd
DL-61	3rd

Invited Speakers:

- CMU-Pose / (4:00pm - 4:15pm)
- G-RMI / (4:15pm - 4:30pm)