Confidence intervals for tracking performance scores

Ricardo Sánchez-Matilla and Andrea Cavallaro

CIS centre for intelligent sensing

Queen Mary University of London

Annotations for multi-object tracking

Tracking evaluation

- $\mathbb{Z}$: Annotation

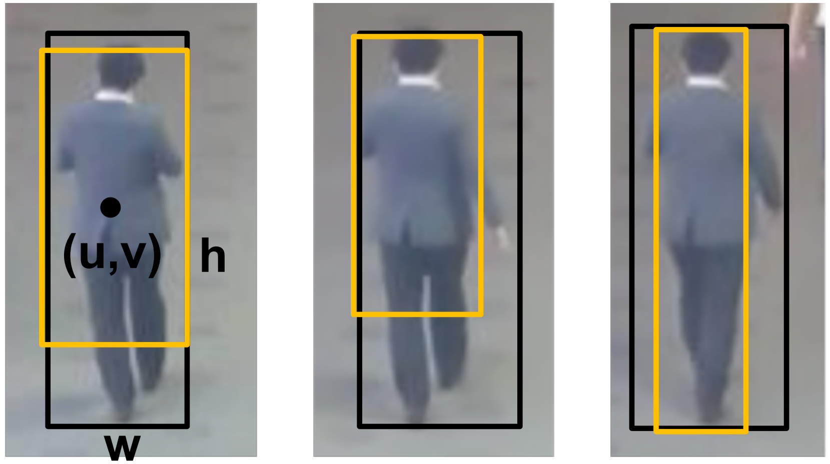
  $\mathbb{Z} = \{\mathbf{z}_k^\lambda : \lambda = 1 \ldots \Lambda; k = 0 \ldots K_\lambda - 1\}$

  $\mathbf{z}_k^\lambda = (u, v, w, h)$

  λ : target identity

  k : time

- $\mathbb{X}$: Estimated object



Tracking score:

$$s = f(\mathbb{Z}, \mathbb{X})$$

Ranking via direct score comparison

MOTB15/16/17

Tracker	Avg Rank	↑MOTA	IDF1	MT	ML	FP	FN	ID Sw.
DH_TRK 1.	16.0	**54.1** ±13.0	49.2	21.6%	**28.4%**	36,196	**216,670**	5,918 (96.1)
HIK_MOT17 2.	11.3	53.9 ±13.7	54.3	**23.7%**	32.0%	27,656	230,042	2,386 (40.3)
NOTBD 3.	18.2	53.9 ±12.7	51.2	21.5%	35.6%	28,912	228,356	2,964 (49.8)
IOUT_Re 4.	18.6	52.7 ±13.0	43.3	20.1%	32.6%	16,529	243,226	6,946 (122.1)
yt_face 5.	15.2	52.6 ±13.1	51.5	23.0%	35.9%	23,894	241,489	2,047 (35.8)

KITTI

	Method	Setting	Code	Out-Noc	Out-All	Avg-Noc	Avg-All
1	M2S_CSPN			1.19 %	1.53 %	0.4 px	0.5 px
2	MS_CSPN			1.25 %	1.61 %	0.4 px	0.5 px
3	DM-Net-Pretrained-30			1.28 %	1.68 %	0.5 px	0.5 px
4	NCA-Net			1.28 %	1.68 %	0.5 px	0.5 px
5	Samsung_System_LSI			1.38 %	1.79 %	0.5 px	0.5 px

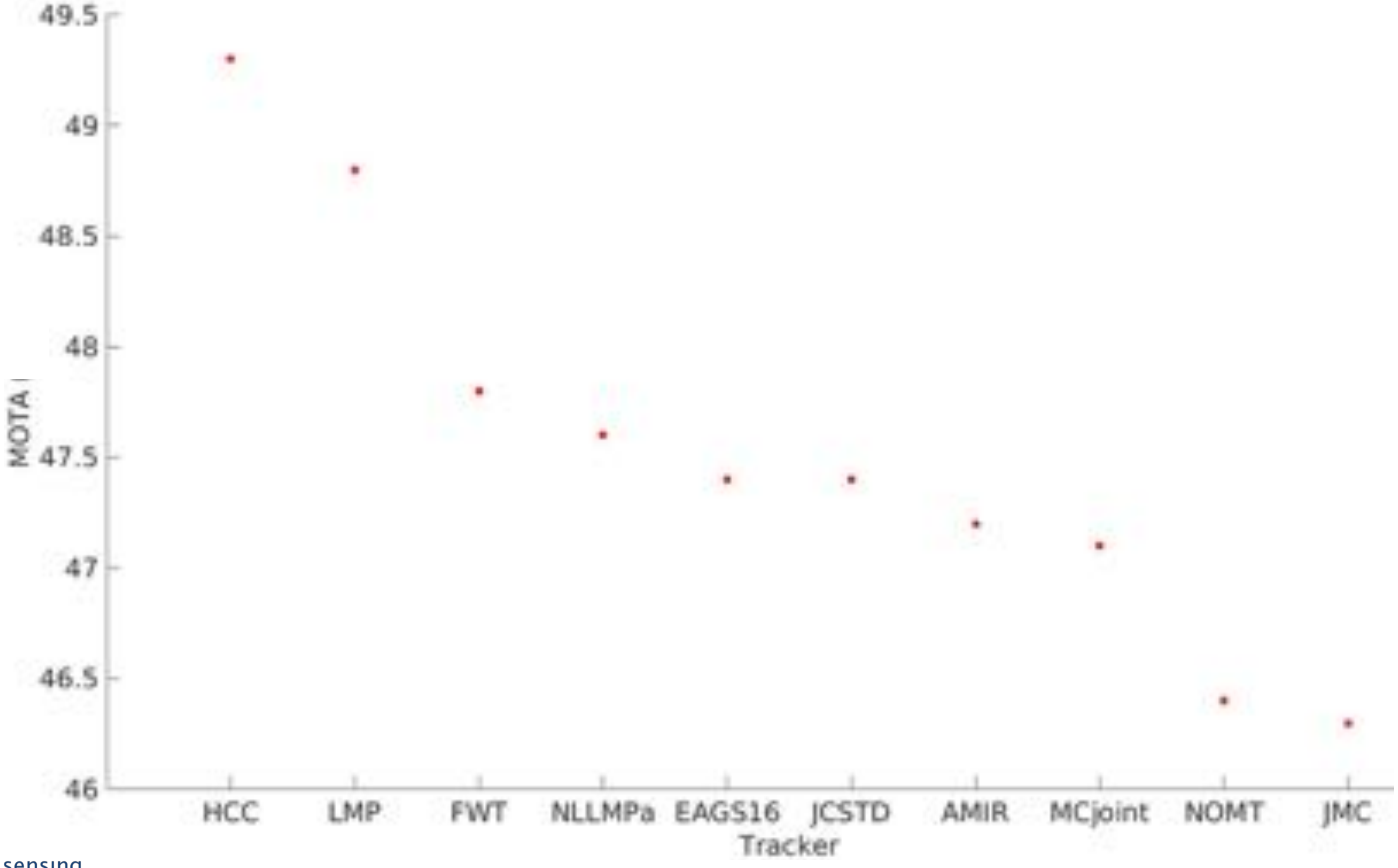
VOT17

		baseline		
	Tracker	**EAO**	**A**	**R**
1.	LSART	0.323 ①	0.493	0.218 ①
2.	CFWCR	0.303 ②	0.484	0.267 ②
3.	CFCF	0.286 ③	0.509	0.281
4.	ECO	0.280	0.483	0.276 ③
5.	Gnet	0.274	0.502	0.276 ③

DukeMTMCT

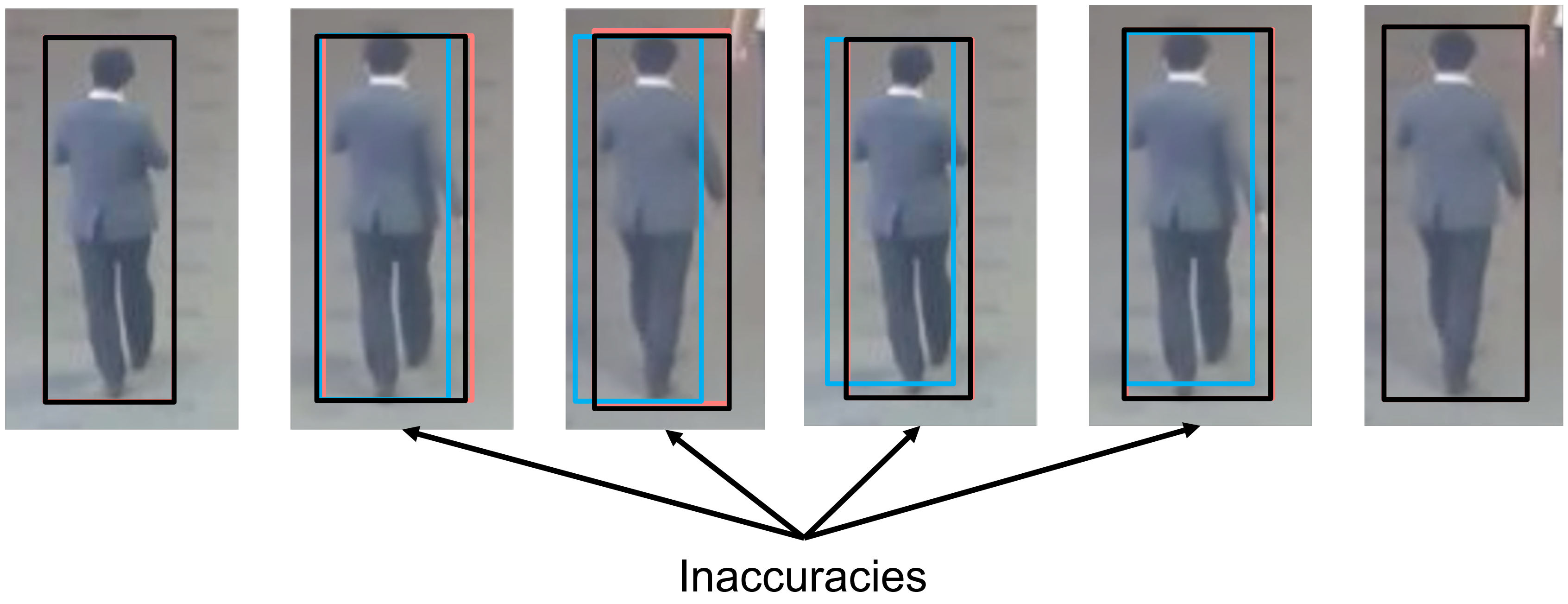
Tracker	↑IDF1	IDP	IDR	MOTA	MOTP	FAF	MT	ML	FP	FN
MTMC_ReID 1.	89.8	92.0	87.7	88.2	79.0	0.05	1,123	17	37,911	86,958
	Z. Zhang, J. Wu, X. Zhang, C. Zhang. Multi-Target, Multi-Camera Tracking by Hierarchical Clustering: Recent Progress on									
DeepCC 2.	89.2	91.7	86.7	87.5	77.1	0.05	1,103	29	37,280	94,399
	E. Ristani, C. Tomasi. Features for Multi-Target Multi-Camera Tracking and Re-									
TAREIDMTMC 3.	83.8	87.6	80.4	83.3	75.5	0.06	1,051	17	44,691	131,220
	N. Jiang, S. Bai, Y. Xu, C. Xing, Z. Zhou, W. Wu. Online Inter-Camera Trajectory Association Exploiting Person Re-Identification and C									
MYTRACKER 4.	80.3	87.3	74.4	78.3	78.4	0.05	914	72	35,580	193,253
	K. Yoon, Y. Song, M. Jeon. Multiple hypothesis tracking algorithm for multi-target multi-camera tracking with disjoint views. I									
MTMC_ReIDp 5.	79.2	89.9	70.7	68.8	77.9	0.07	726	143	52,408	277,762
	Z. Zhang, J. Wu, X. Zhang, C. Zhang. Multi-Target, Multi-Camera Tracking by Hierarchical Clustering: Recent Progress on									

Ranking via direct score comparison in MOTB16 [1]



[1] https://motchallenge.net

Annotation procedures



Annotations

Manual annotation

Linearly interpolated annotation

Annotations in tracking datasets

Dataset	Linear Interpolation
CAVIAR	×
TUD	√
ETH	√
PETS09	√
KITTI	Not available
i-LIDS	√
MOTB15	√
MOTB16	√
MOTB17	√

Annotations in tracking

- Performance evaluation uses manual annotations

- Manual annotation process [2]
  - Tedious
  - Expensive
  - (Potentially) unfeasible

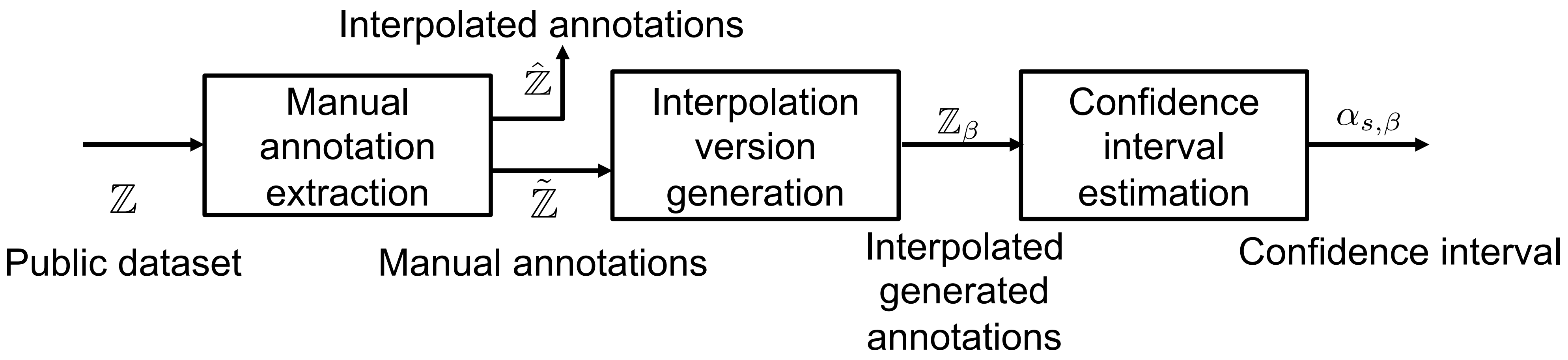
- Semi-automatic annotation process
  - Less tedious and less expensive
  - (Potentially) Inaccurate → inaccurate evaluations

- This work: accounts for inaccuracies due to linear interpolation

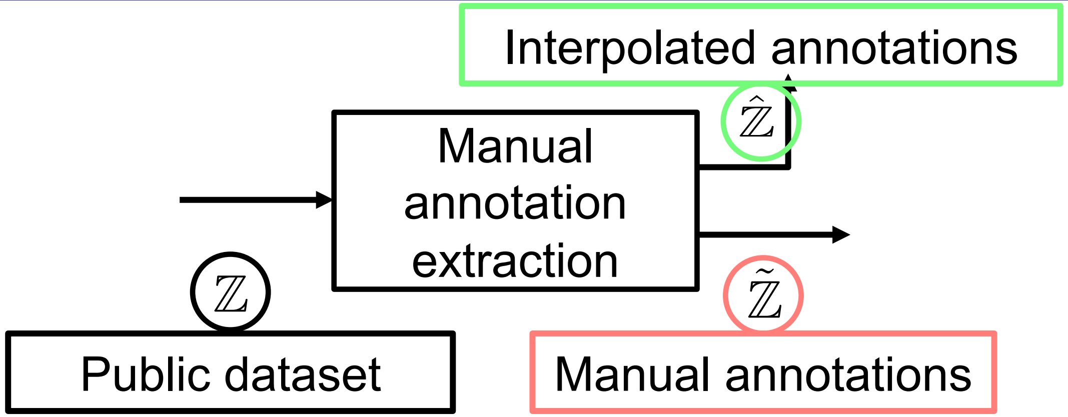
[2] Vondrick, Patterson and Ramanan, "Efficiently scaling up crowdsourced video annotations", IJCV 2013

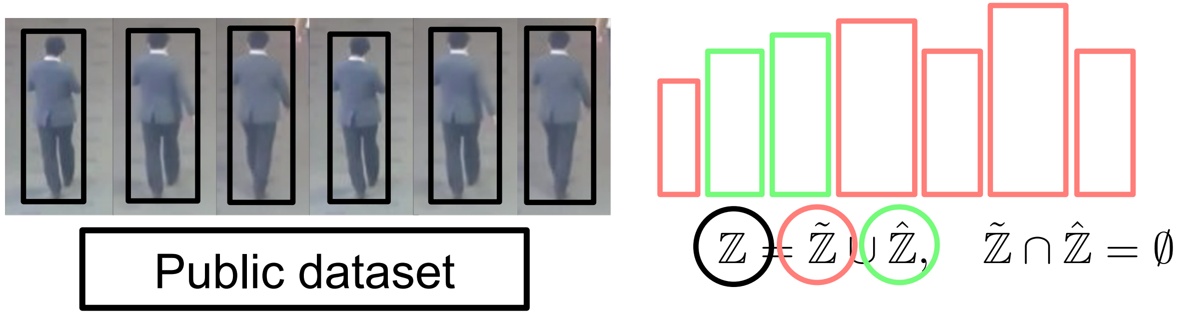
How to account for annotation inaccuracies?

- We propose to estimate <u>confidence intervals</u> for a given dataset:
  - With **unknown** annotation procedure
  - **Without** further annotations



Manual annotation extraction





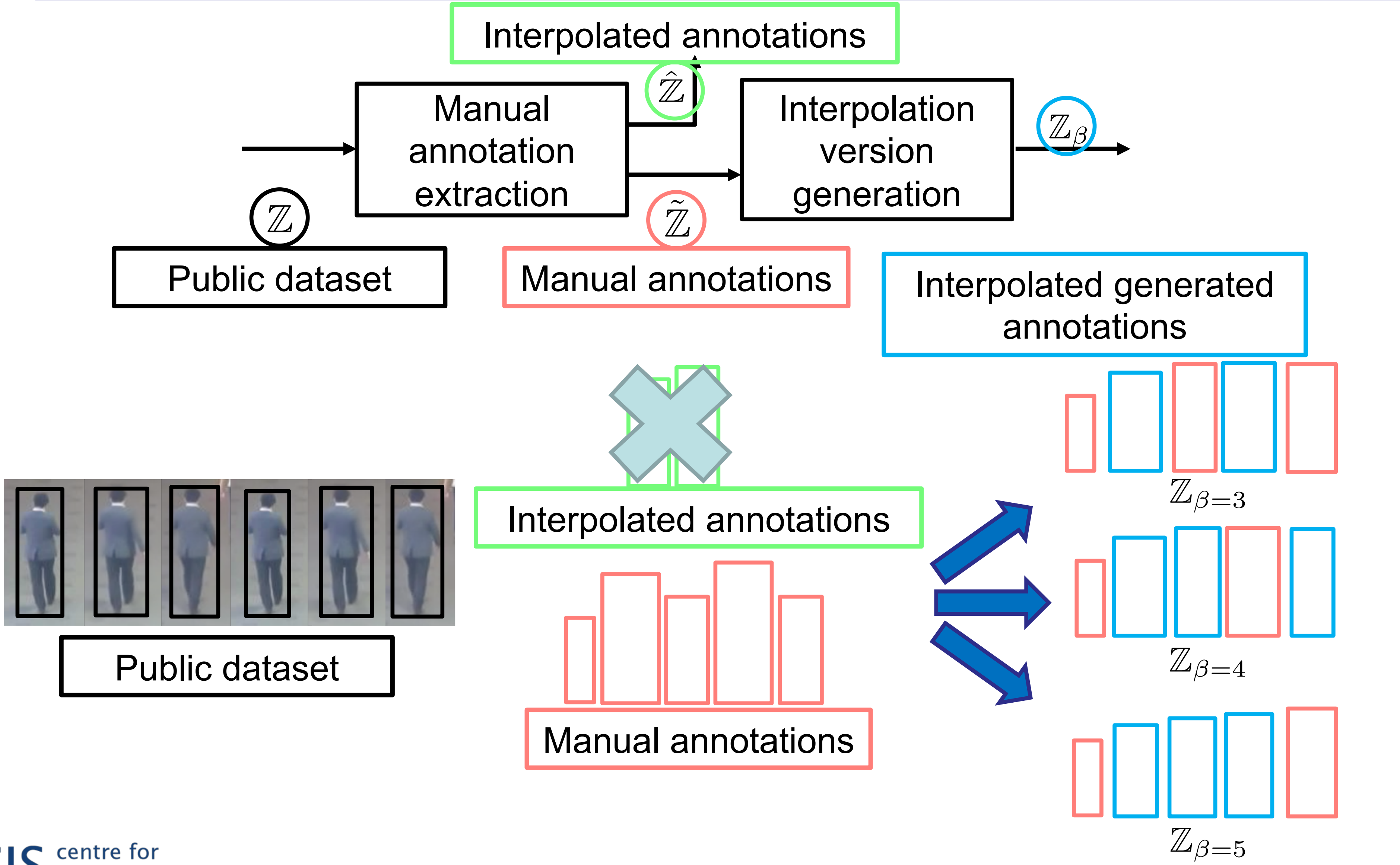
- Linear interpolation detection:

$$\tilde{\mathbb{Z}}^{\lambda} = \{\mathbf{z}_k^{\lambda} : \mathbf{z}_{uu}^{\lambda}, \mathbf{z}_{vv}^{\lambda}, \mathbf{z}_{ww}^{\lambda}, \mathbf{z}_{hh}^{\lambda} \neq 0\}$$

λ : target identity

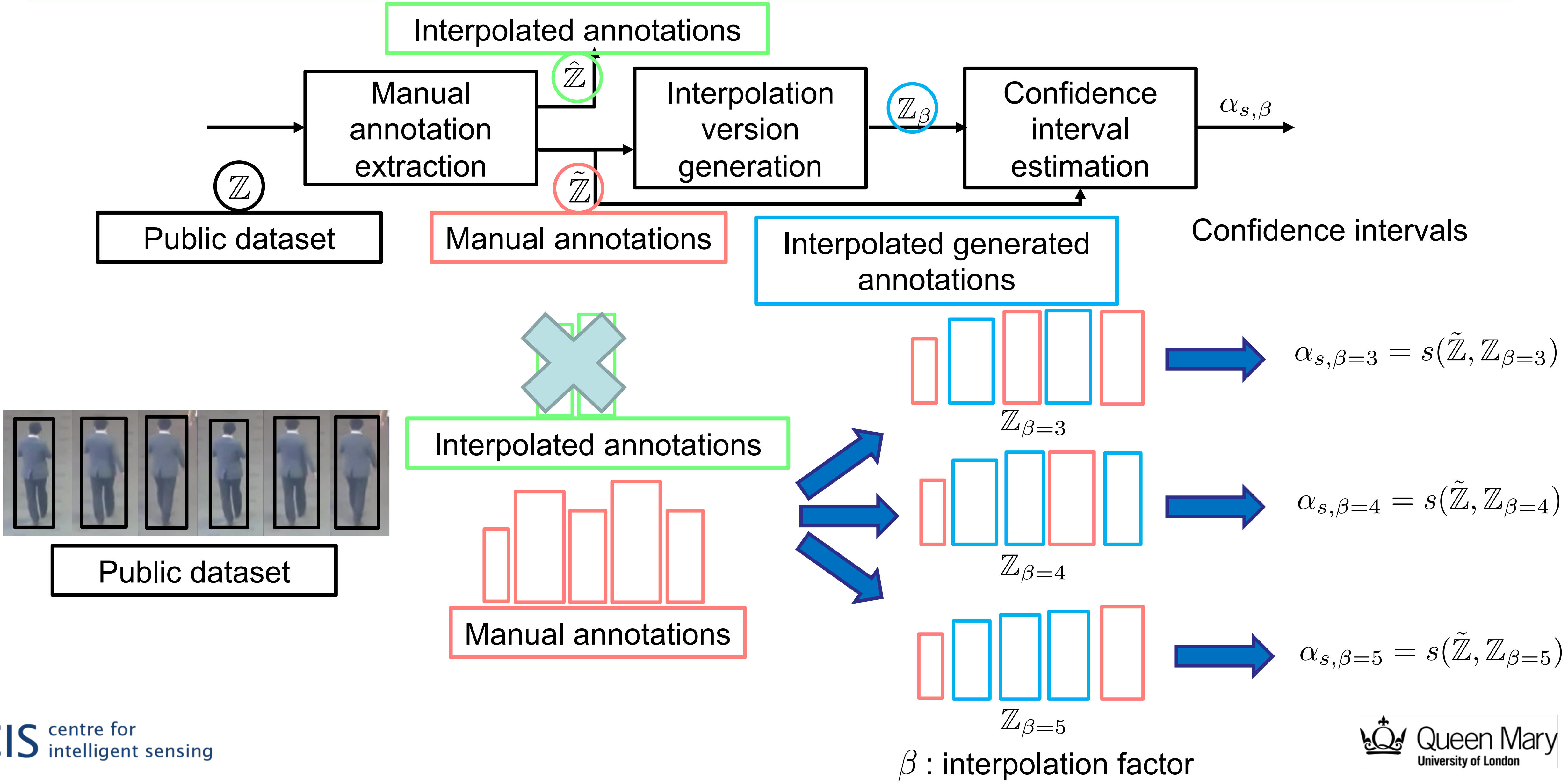
$\mathbf{z}_{\cdot\cdot}^{\lambda}$: second derivative on $\cdot\cdot$ component

Interpolation version generation



β : interpolation factor

Confidence interval estimation



Experiment setup

- Benchmark dataset: MOTB16 [2]
- Interpolation factor: $\beta \in \{3, 6, 9, 12\}$

- Detected percentage of annotations with linear interpolation: 39.7% ≈ $\beta = 3$

Confidence interval

$$\alpha_{s,\beta} = s(\tilde{\mathbb{Z}}, \mathbb{Z}_\beta)$$

$\tilde{\mathbb{Z}}$: manually annotated subset

$\mathbb{Z}_\beta$: linearly interpolated

$$s(\cdot,\cdot) = 100 - MOTA \qquad MOTA = 1 - \frac{1}{N}\sum_{\lambda=0}^{\Lambda}\sum_{k=0}^{\tilde{K}_\lambda - 1}(FN_k^\lambda + FP_k^\lambda + IDSW_k^\lambda)$$

$$s(\cdot,\cdot) = 100 - MOTP \qquad MOTP = \frac{1}{N}\sum_{\lambda=0}^{\Lambda}\sum_{k=0}^{\tilde{K}'_\lambda - 1}\frac{\tilde{\mathbf{z}}_k^\lambda \cap \mathbf{z}_{k,\beta}^\lambda}{\tilde{\mathbf{z}}_k^\lambda \cup \mathbf{z}_{k,\beta}^\lambda}$$

N: # annotations

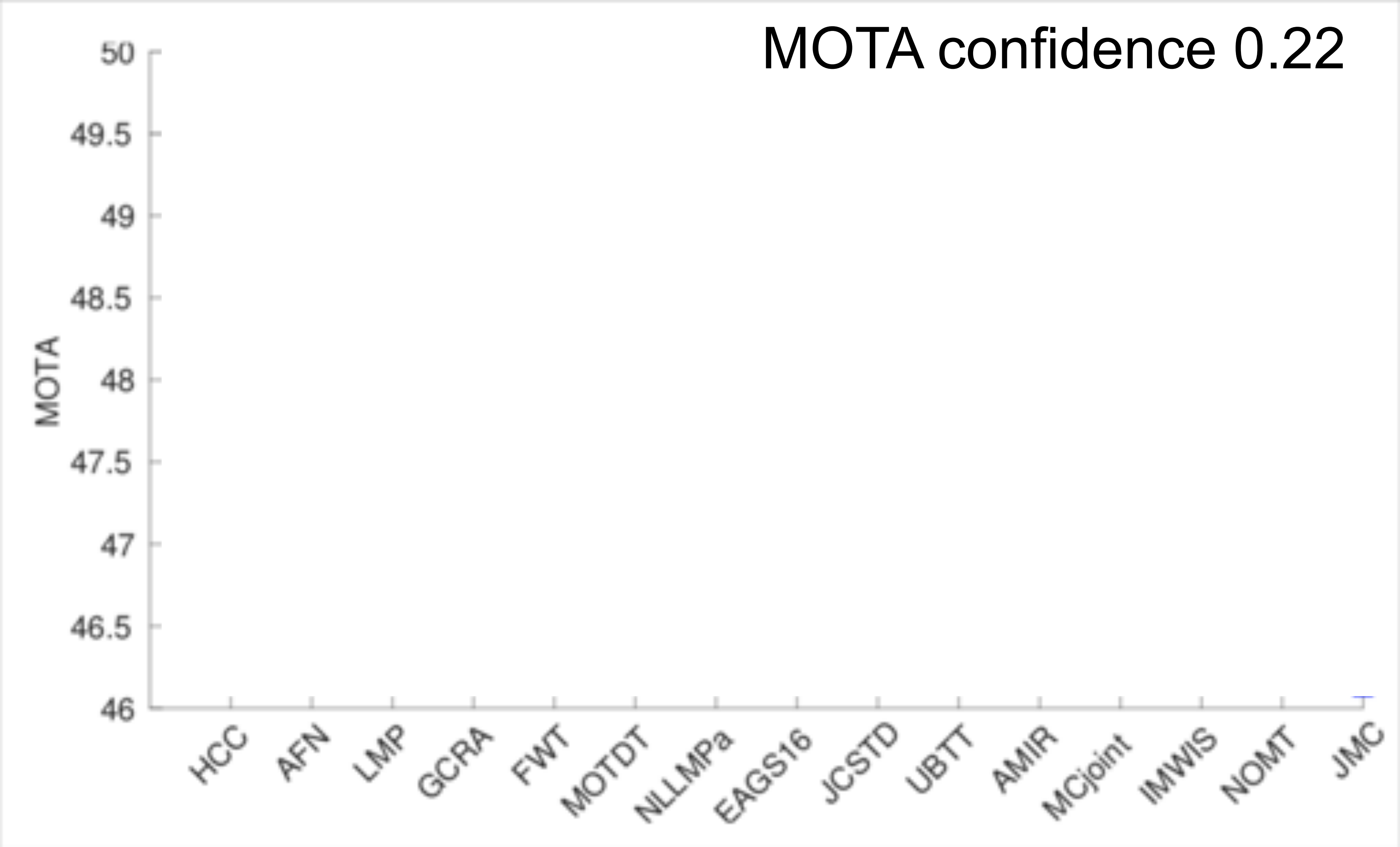
β: interpolation factor

CIS centre for intelligent sensing

[2] https://motchallenge.net

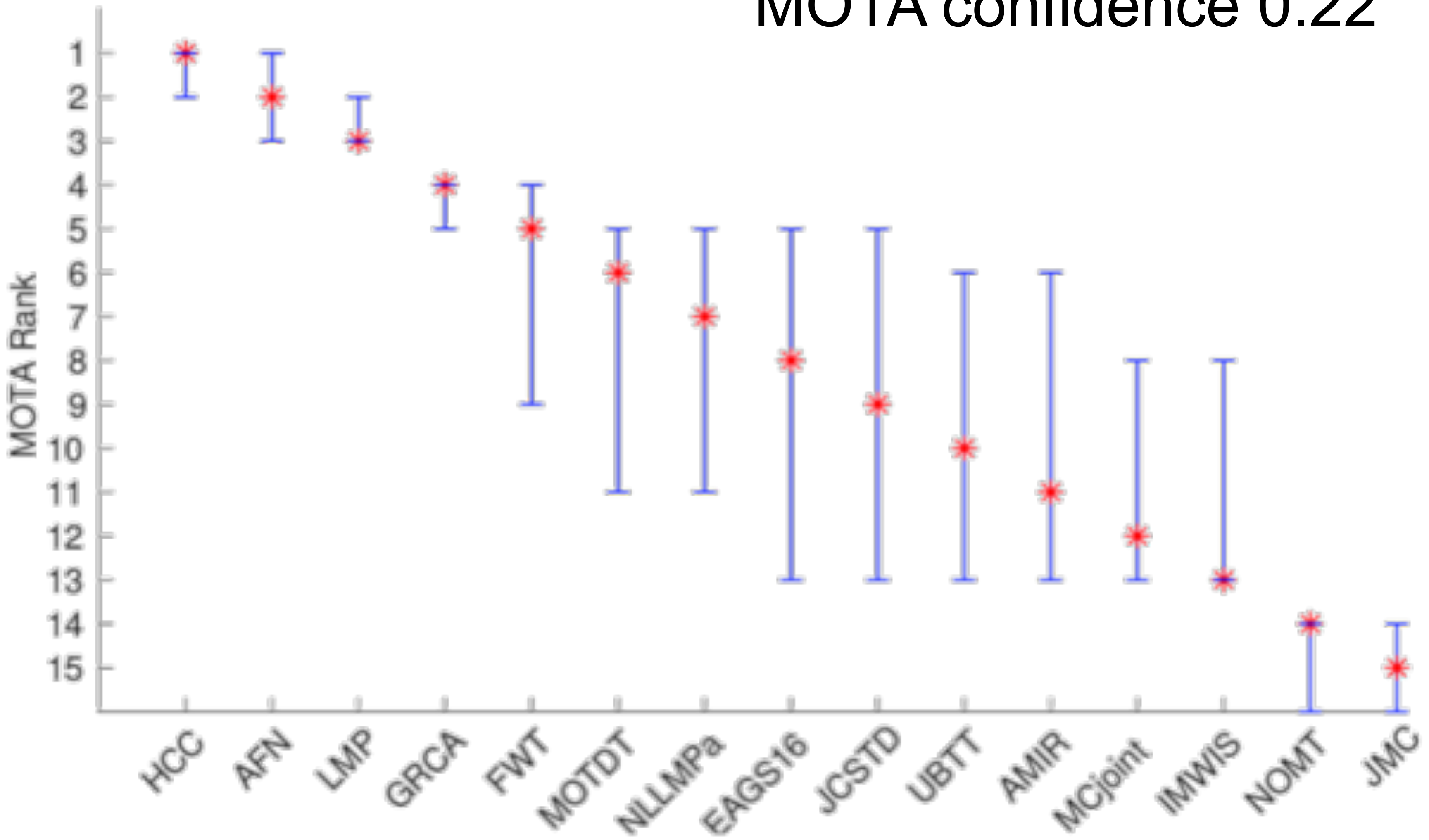
Queen Mary University of London

Impact on MOTB16 – MOTA



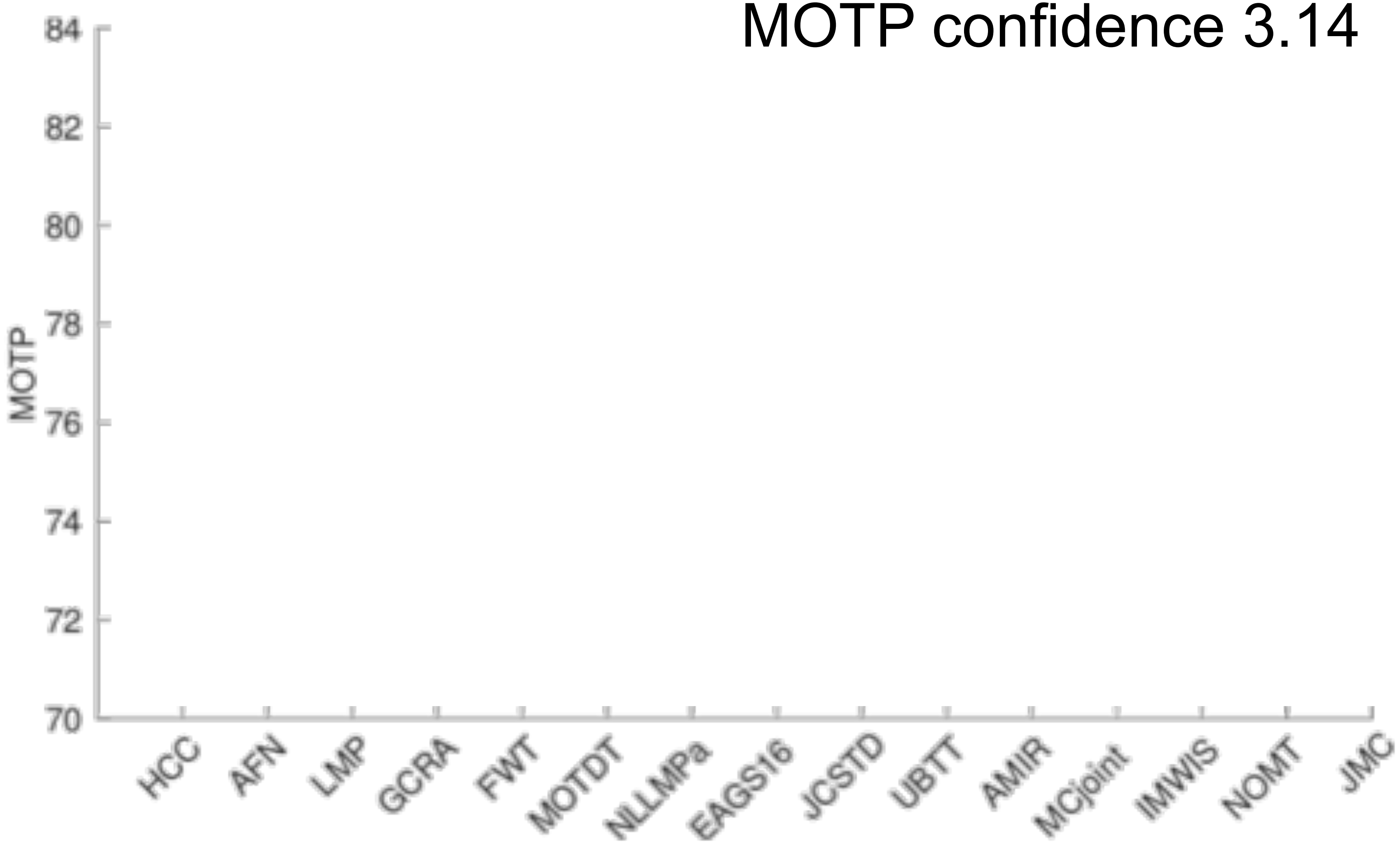
Impact on MOTB16 – MOTA ranking

MOTA confidence 0.22

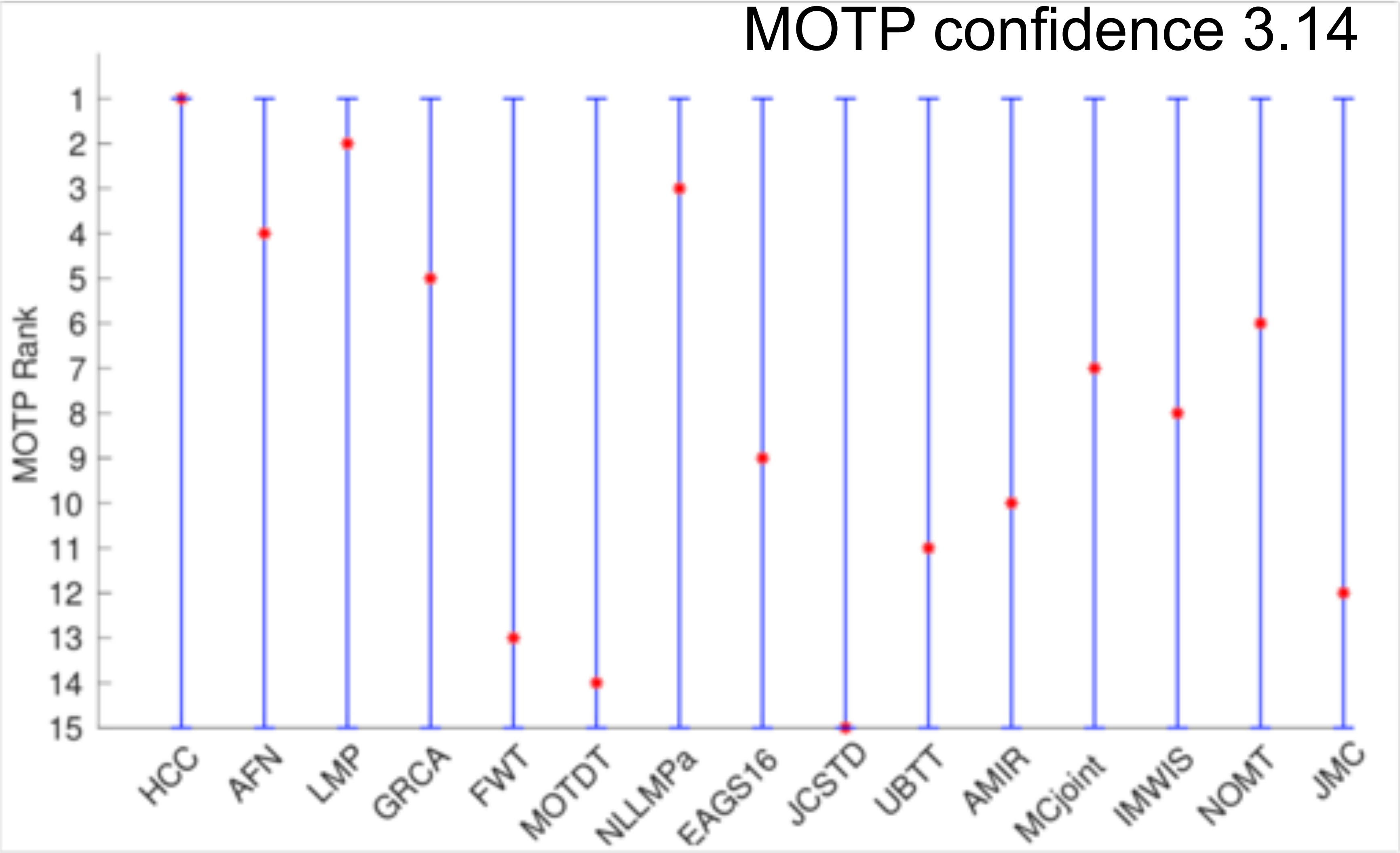


Impact on MOTB16 – MOTP

MOTP confidence 3.14



Impact on MOTB16 – MOTP ranking



Conclusion

- Interpolation
  - large scale datasets
  - inaccuracies

- Confidence intervals
  - consider annotation inaccuracies from an already annotated dataset
  - no need of further annotations

- Future work
  - consider other type of semi-automatic annotation (e.g. tracking)
  - consider other applications (e.g. DNN tasks)

Paper