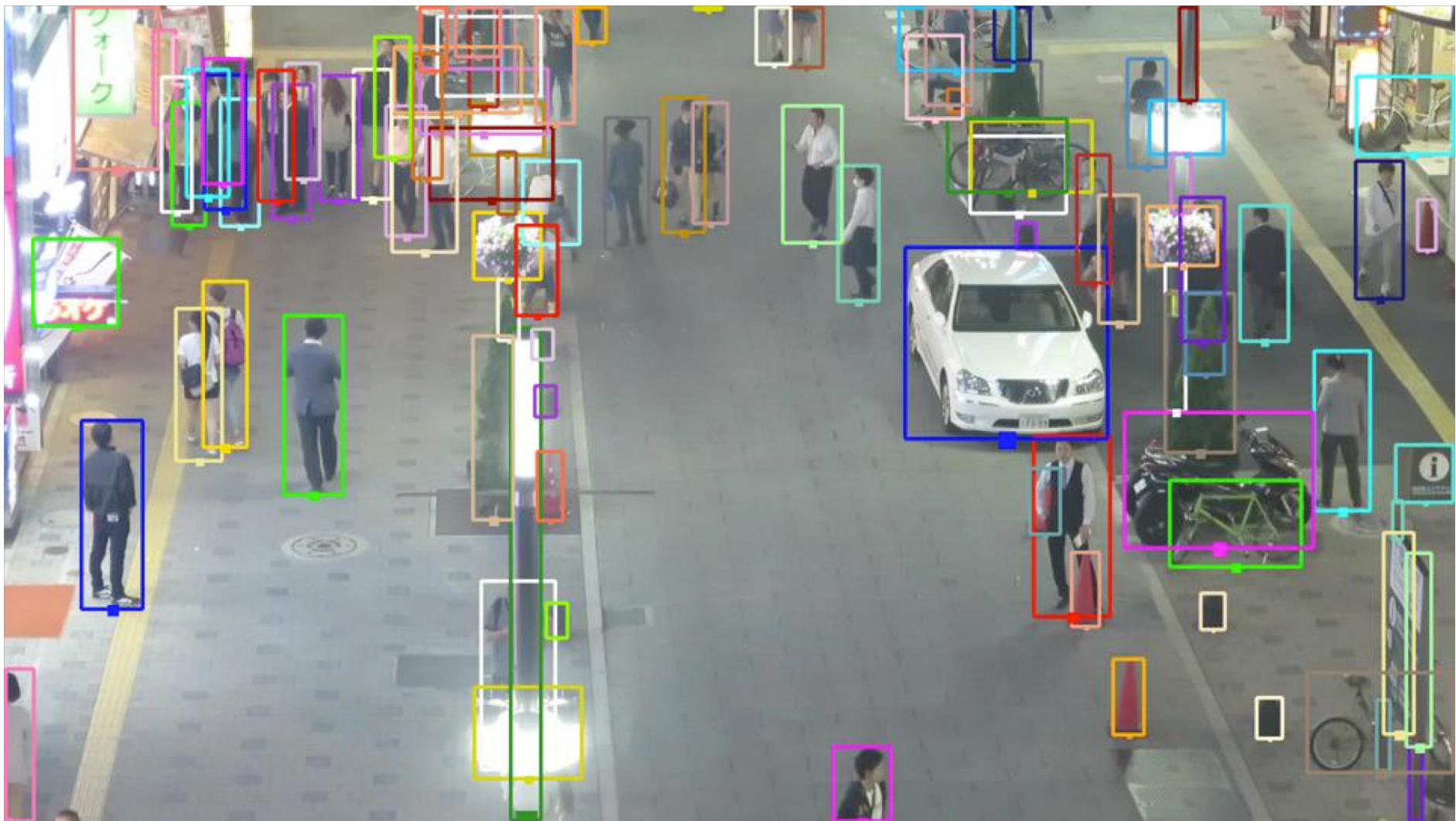


Confidence intervals for tracking performance scores

Ricardo Sánchez-Matilla and Andrea Cavallaro

Annotations for multi-object tracking



Tracking evaluation

□ \mathbb{Z} : Annotation

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

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

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	16.0	54.1 ± 13.0	49.2	21.6%	28.4%	36,196	216,670	5,918 (96.1)
1.								
HIK MOT17	11.3	53.9 ± 13.7	54.3	23.7%	32.0%	27,656	230,042	2,386 (40.3)
2.								
NOTBD	18.2	53.9 ± 12.7	51.2	21.5%	35.6%	28,912	228,356	2,964 (49.8)
3.								
IOUT_Re	18.6	52.7 ± 13.0	43.3	20.1%	32.6%	16,529	243,226	6,946 (122.1)
4.								
yt_face	15.2	52.6 ± 13.1	51.5	23.0%	35.9%	23,894	241,489	2,047 (35.8)
5.								

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

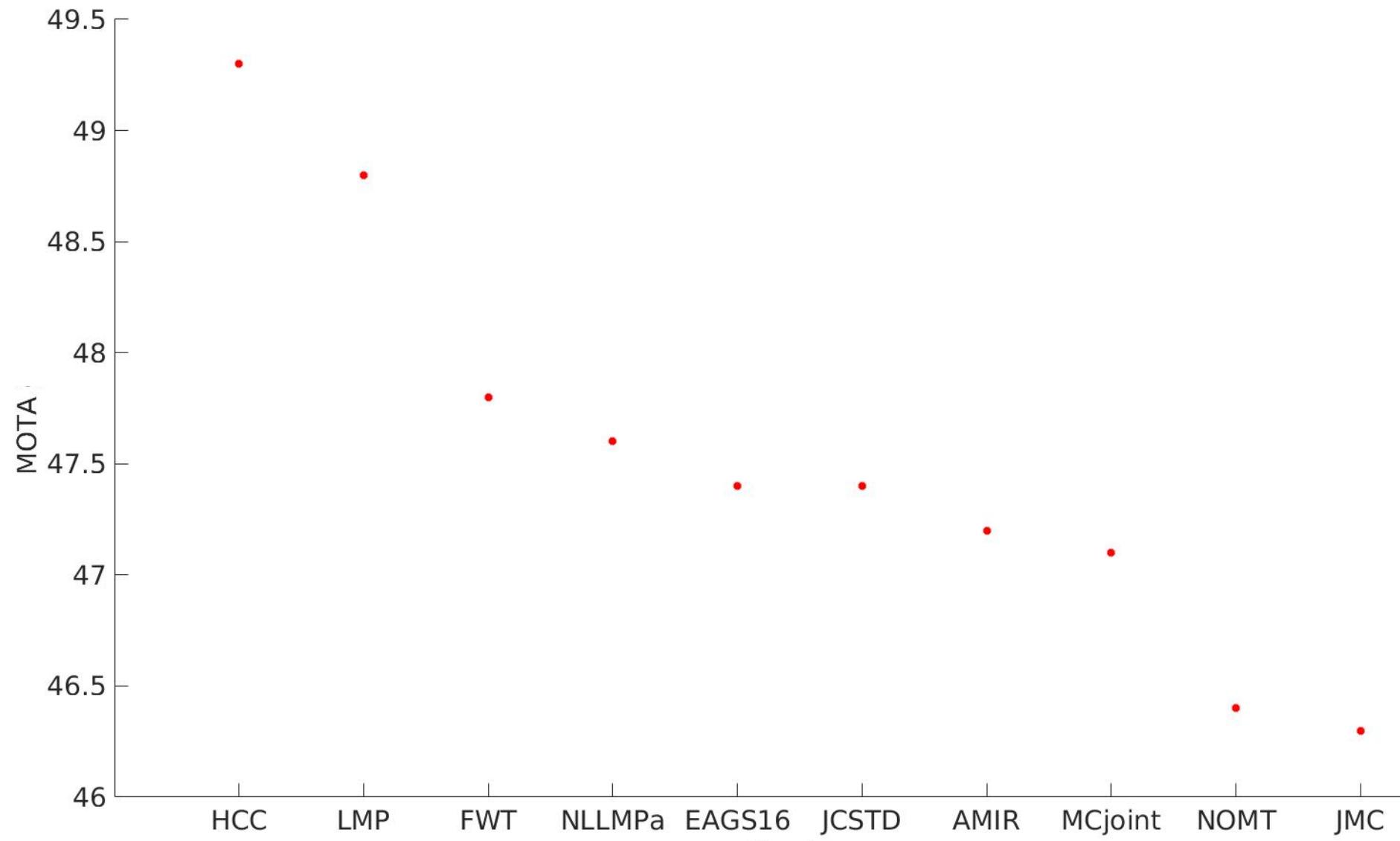
VOT17

	Tracker	baseline		
		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

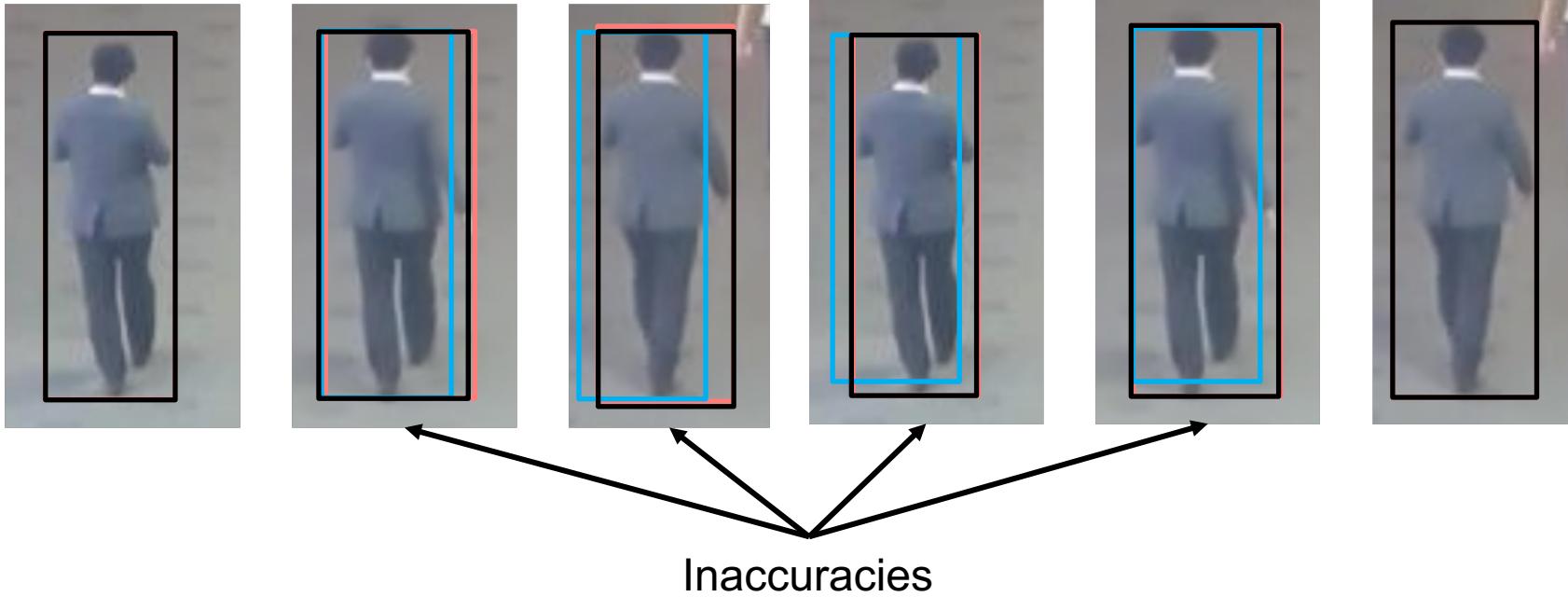
DukeMTMCT

Tracker	↑IDF1	IDP	IDR	MOTA	MOTP	FAF	MT	ML	FP	FN
MTMC_RelID	89.8	92.0	87.7	88.2	79.0	0.05	1,123	17	37,911	86,958
1.										
DeepCC	89.2	91.7	86.7	87.5	77.1	0.05	1,103	29	37,280	94,399
2.										
TAREIDMTMC	83.8	87.6	80.4	83.3	75.5	0.06	1,051	17	44,691	131,220
3.										
MYTRACKER	80.3	87.3	74.4	78.3	78.4	0.05	914	72	35,580	193,253
4.										
MTMC_RelDp	79.2	89.9	70.7	68.8	77.9	0.07	726	143	52,408	277,762
5.										

Ranking via direct score comparison in MOTB16 [1]



Annotation procedures



- Annotations
- Manual annotation
- Linearly interpolated annotation

Annotations in tracking datasets

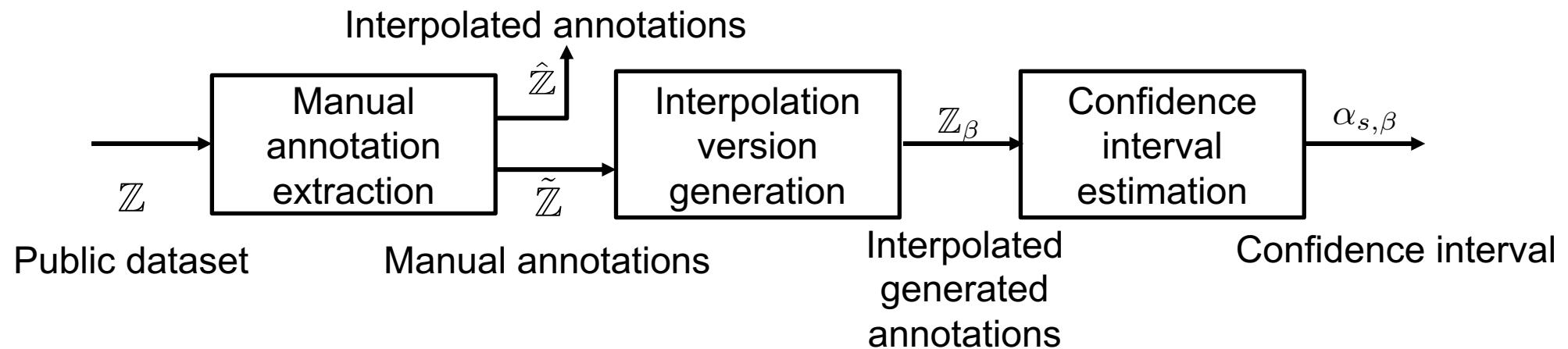
Dataset	Linear Interpolation
CAVIAR	✗
TUD	✓
ETH	✓
PETS09	✓
KITTI	Not available
i-LIDS	✓
MOTB15	✓
MOTB16	✓
MOTB17	✓

Annotations in tracking

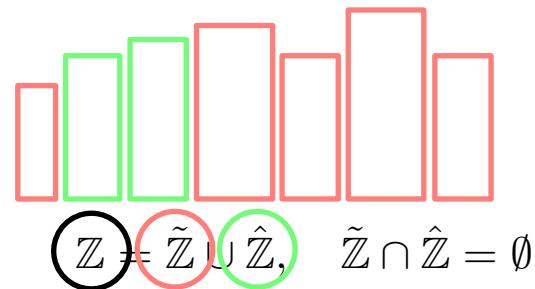
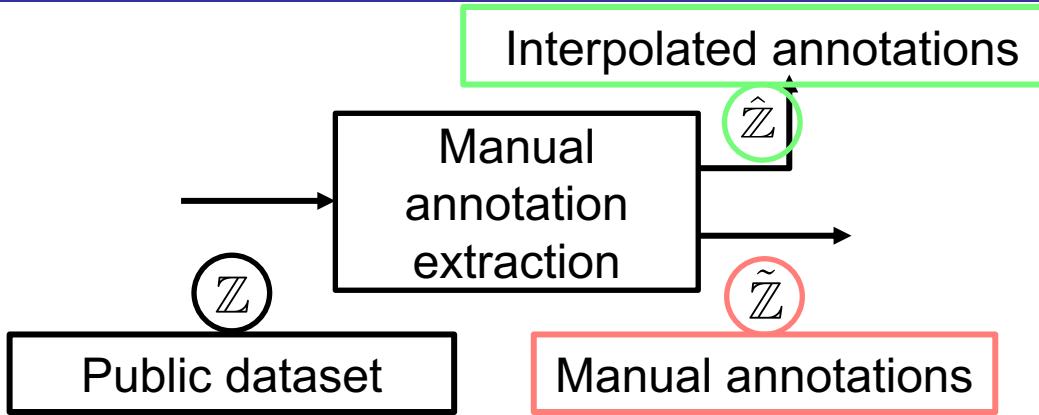
- Performance evaluation uses manual annotations
- Manual annotation process [2]
 - Tedious
 - Expensive
 - (Potentially) unfeasible
- Semi-automatic annotation process
 - Less tedious and less expensive
 - (Potentially) Inaccurate → inaccurate evaluations
- This work: accounts for inaccuracies due to linear interpolation

How to account for annotation inaccuracies?

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



Manual annotation extraction



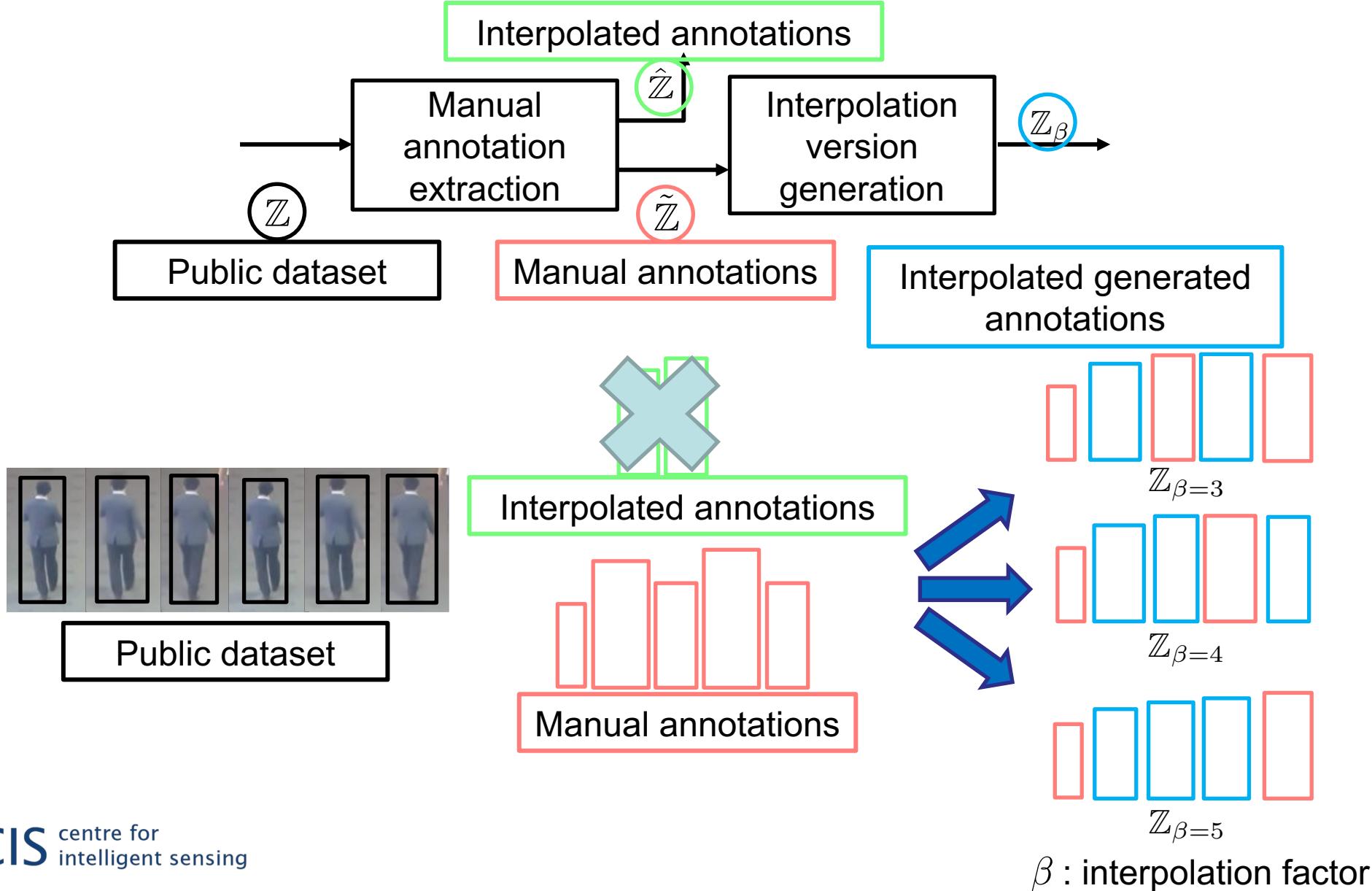
- Linear interpolation detection:

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

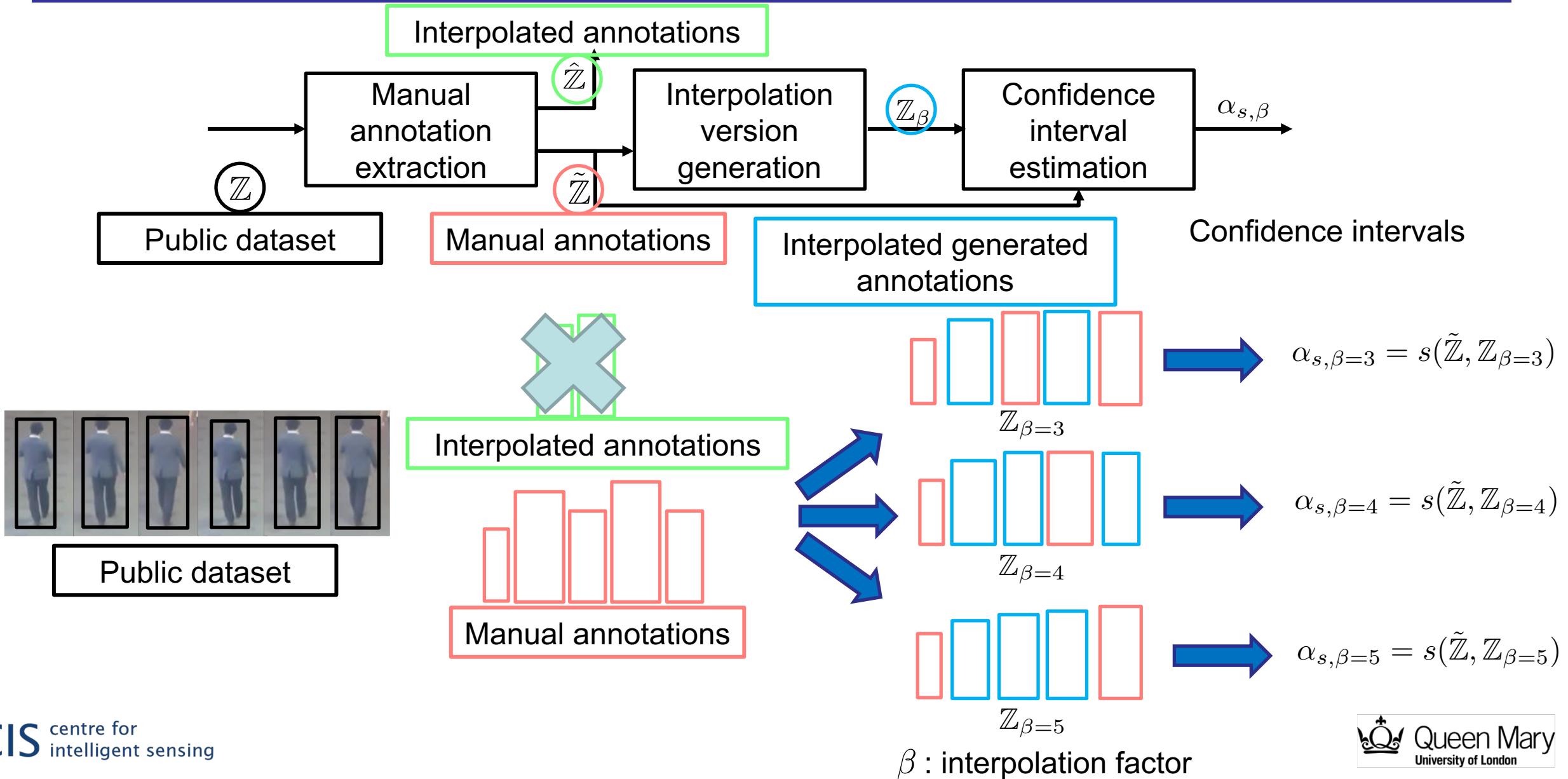
λ : target identity

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

Interpolation version generation



Confidence interval estimation



Experiment setup

- Benchmark dataset: MOTB16 [2]
- Interpolation factor: $\beta \in \{3, 6, 9, 12\}$
- Detected percentage of annotations with linear interpolation: $39.7\% \approx \beta = 3$

Confidence interval

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

$$s(\cdot, \cdot) = 100 - MOTA$$

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

\mathbb{Z}_β : linearly interpolated

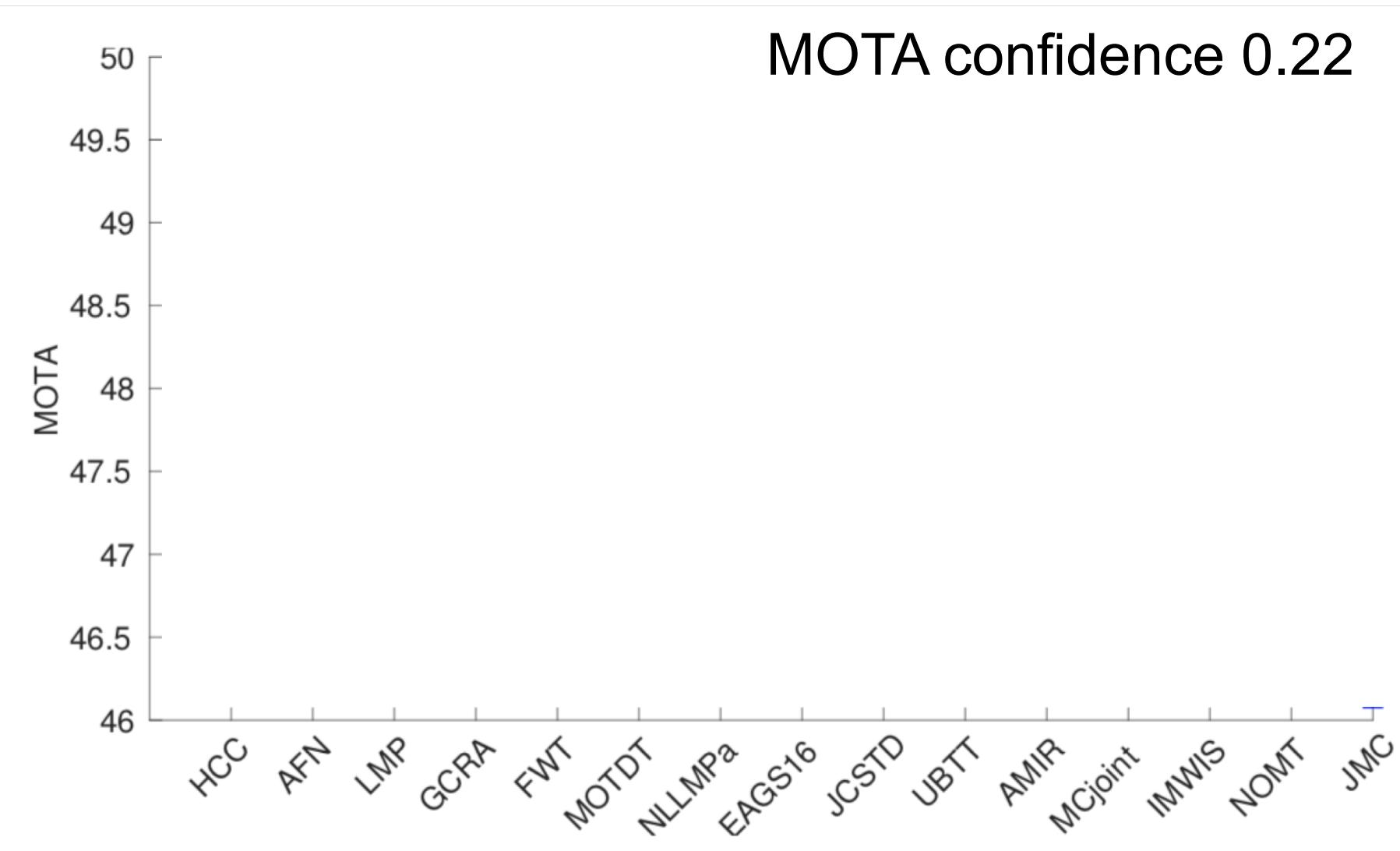
$$s(\cdot, \cdot) = 100 - MOTP$$

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

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

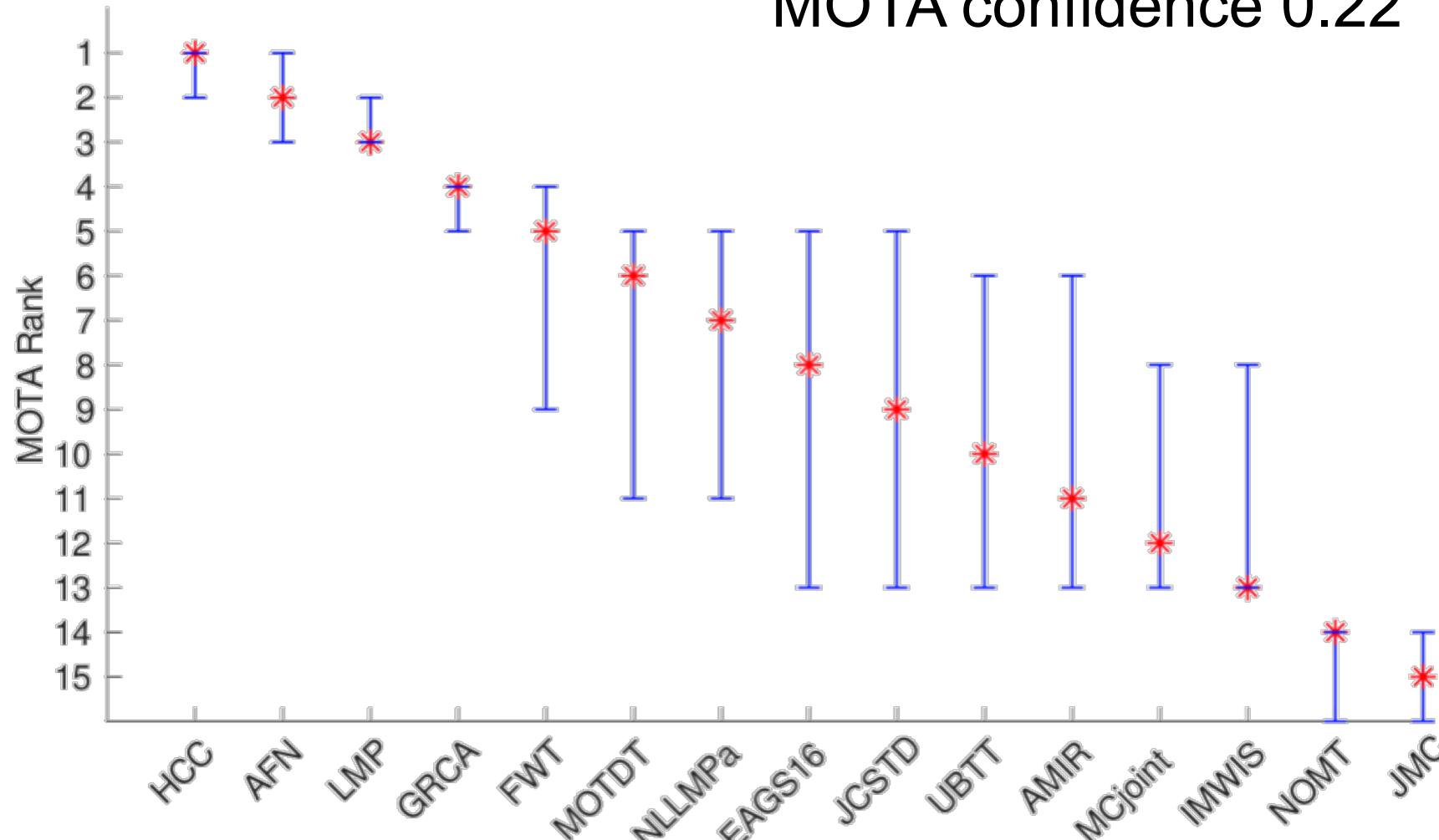
N : # annotations
 β : interpolation factor

Impact on MOTB16 – MOTA

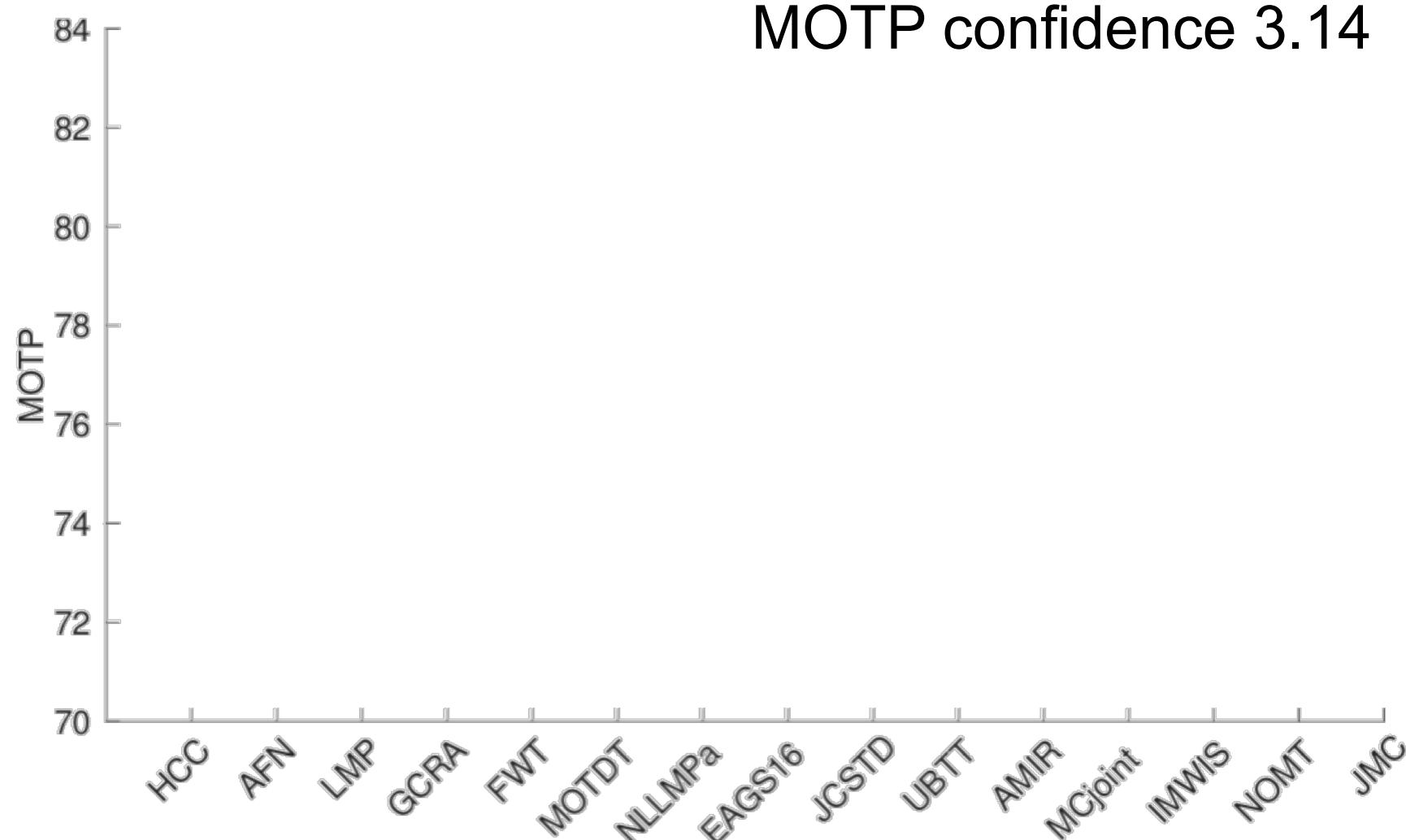


Impact on MOTB16 – MOTA ranking

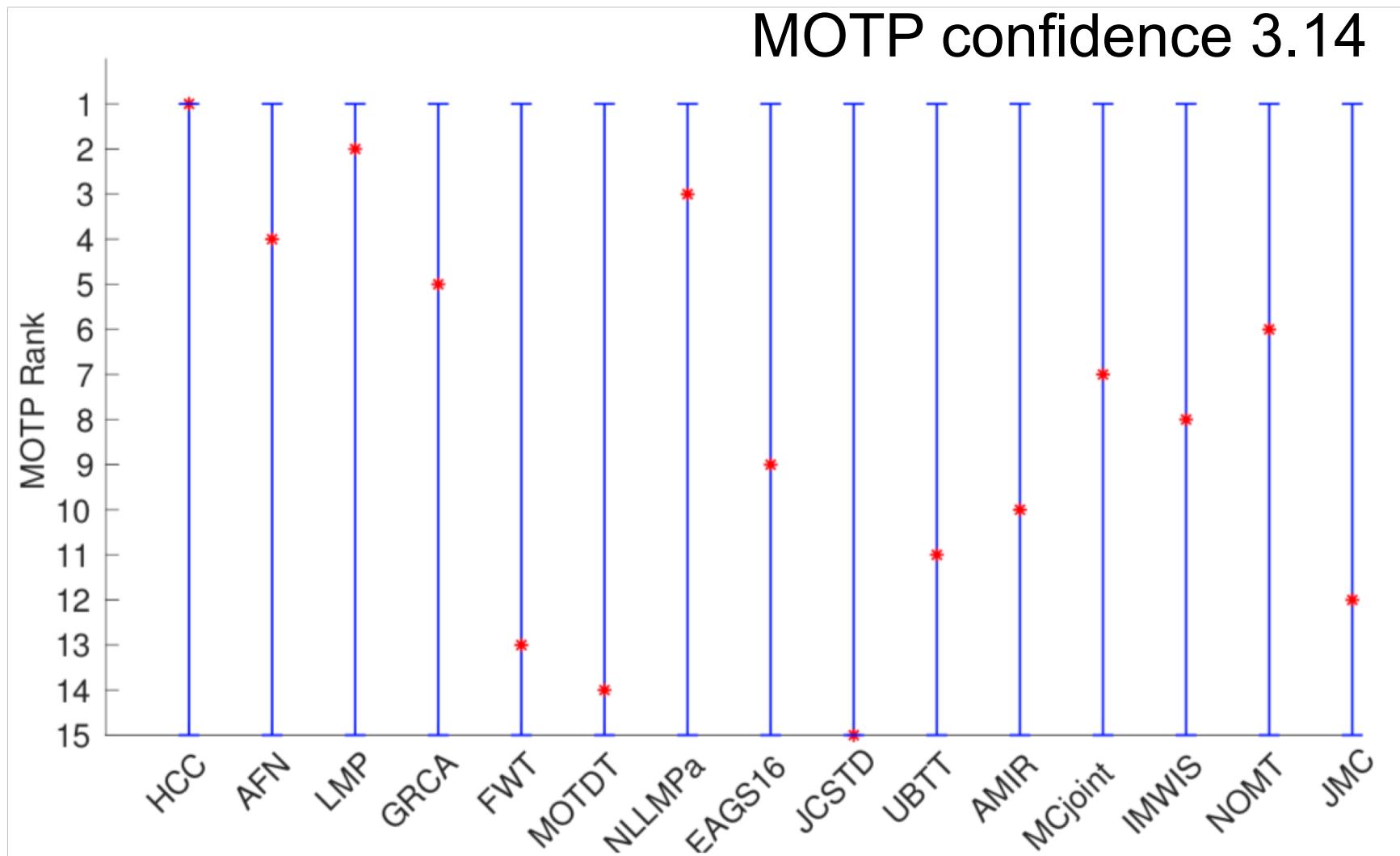
MOTA confidence 0.22



Impact on MOTB16 – MOTP



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