Confidence intervals for tracking performance scores

Ricardo Sánchez-Matilla and Andrea Cavallaro

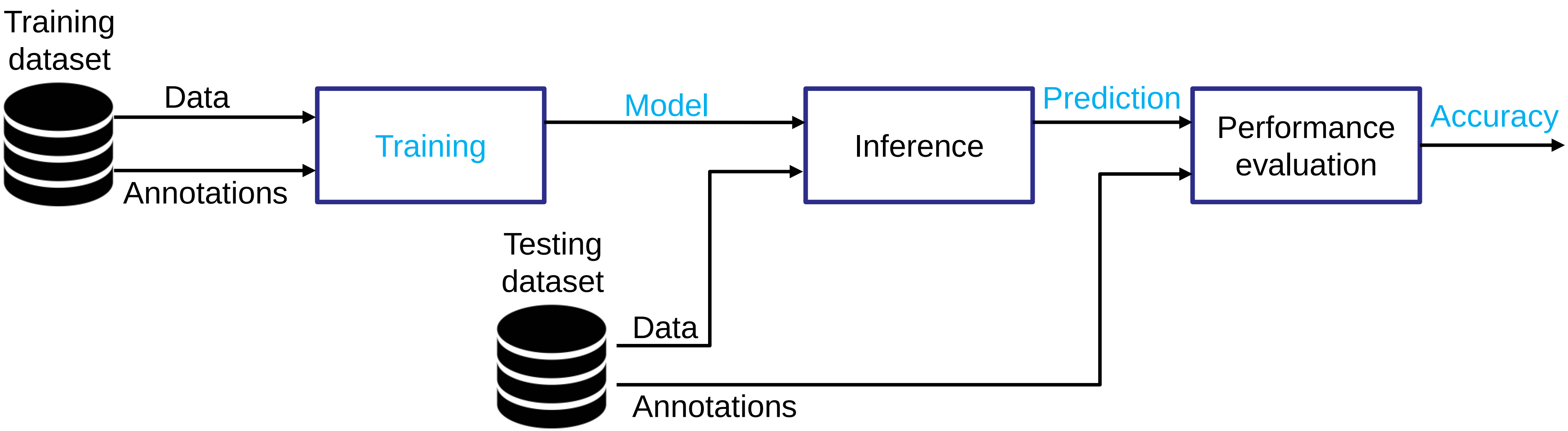
Published in: *IEEE International Conference on Image Processing (ICIP) 2018*

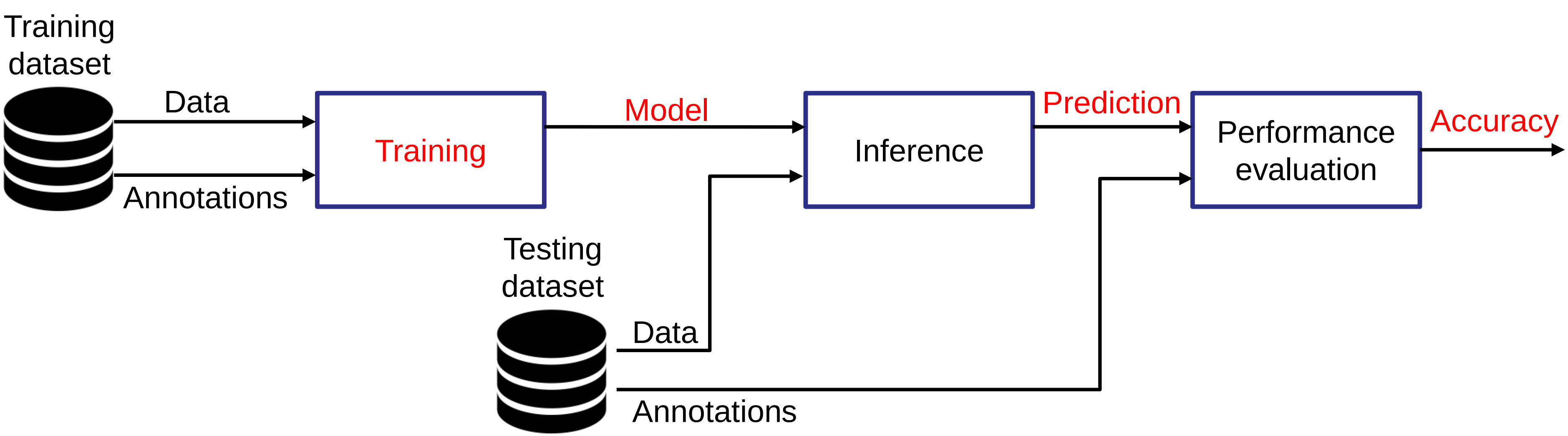
Centre for Intelligent Sensing
Queen Mary University of London

CIS centre for intelligent sensing

Queen Mary University of London

Performance evaluation (ranking)



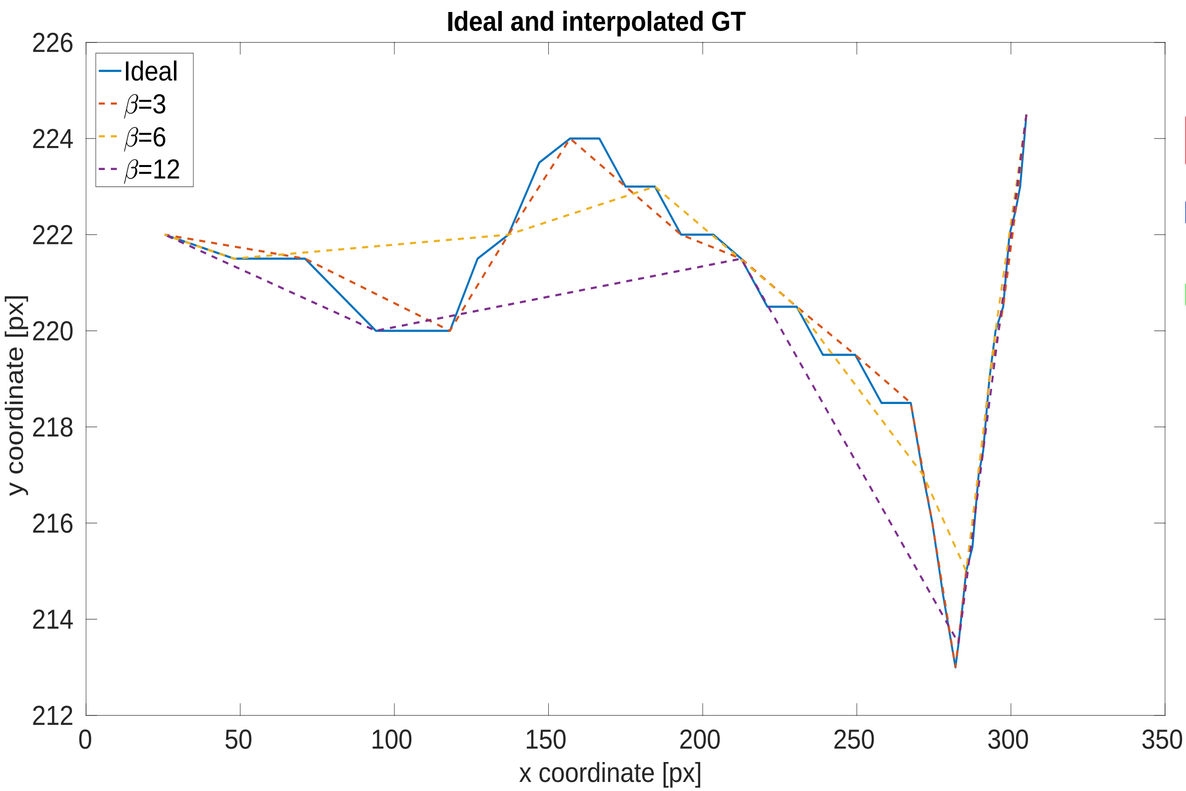


Annotations

[1] Vondrick, Patterson and Ramanan, “Efficiently scaling up crowdsourced video annotations”, IJCV 2013

- Performance evaluation requires manual annotations
- Manual annotations are [1]:
  - Tedious
  - Expensive (time and economically wise)
  - (Potentially) inaccurate
  - (Potentially) unfeasible to retrieve
- Inaccurate annotations produce inaccurate evaluations

Tracking: Interpolated annotations



Manual (ideal) annotation

Interpolated annotation

Publicly available annotation (uses linear interpolation)

β: interpolation factor

Tracking: Datasets

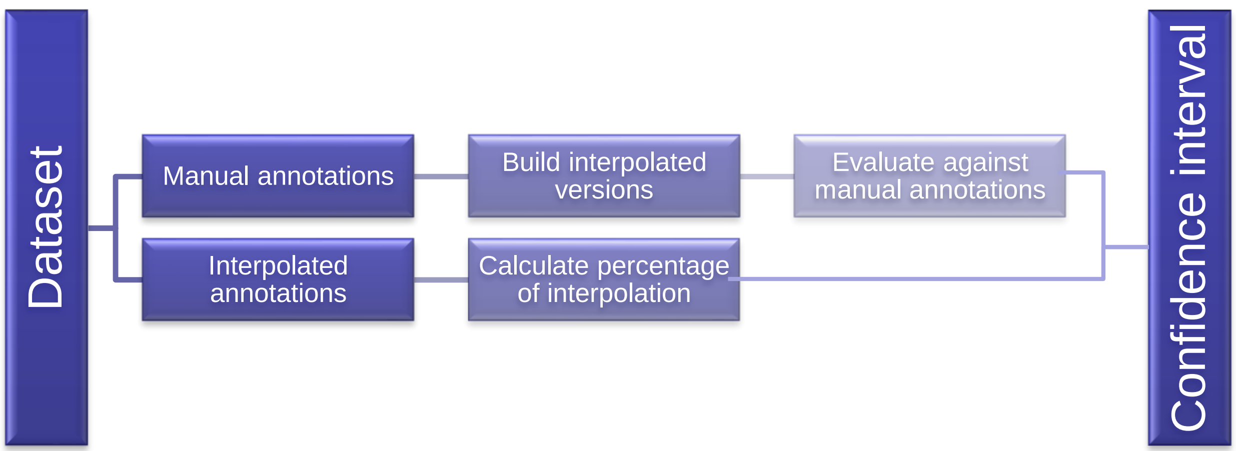
- Most public datasets use linear interpolation for their annotations

Dataset	Tool	Linear Interpolation
CAVIAR	CaviarGui	
TUD	NA	✓
ETH	NA	✓
PETS09	NA	✓
KITTI	MT	NA
i-LIDS	ViPER	✓
MOTB15	VATIC	✓
MOTB16	NA	✓

NA: not available

Confidence intervals

- Annotation inaccuracies should be taken into account
- Estimate uncertainty in annotations for a given dataset:
  - Unknown interpolation
  - Without doing further annotations
  - Estimate confidence interval



Separating manual/interpolated annotations

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

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

λ: target identity

k: time

$$\mathbb{Z} = \tilde{\mathbb{Z}} \cup \hat{\mathbb{Z}}, \quad \tilde{\mathbb{Z}} \cap \hat{\mathbb{Z}} = \emptyset$$

$\mathbb{Z}$: public dataset

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

$\hat{\mathbb{Z}}$: linearly interpolated subset ← Second derivatives equal to zero

- Percentage of interpolation in the dataset: $\frac{|\hat{\mathbb{Z}}|}{|\mathbb{Z}|}$

Building interpolated versions

- Generate interpolated versions of the manual subset through a decimation-interpolation procedure

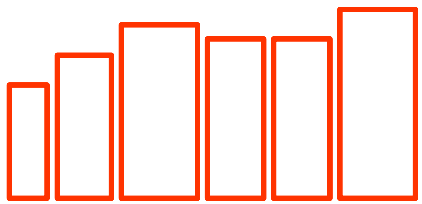
$$\mathbb{Z} \longrightarrow \begin{cases} \tilde{\mathbb{Z}} \longrightarrow \mathbb{Z}_\beta \\ \hat{\mathbb{Z}} \end{cases}$$

$\beta \in \{3, 6, 9, 12\}:$ interpolation factor

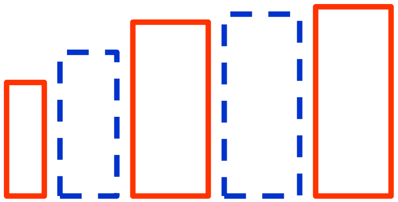
$\mathbb{Z}$: public dataset

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

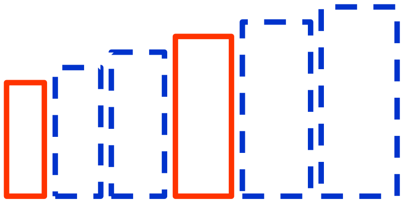
$\hat{\mathbb{Z}}$: linearly interpolated subset



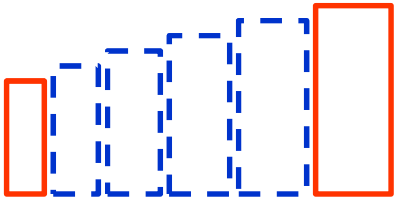
$\tilde{\mathbb{Z}}$



$\mathbb{Z}_{\beta=3}$



$\mathbb{Z}_{\beta=4}$



$\mathbb{Z}_{\beta=5}$

Confidence intervals

- Compare interpolated version against manual version

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

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

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

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

$\mathbb{Z}$: public dataset

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

$\hat{\mathbb{Z}}$: linearly interpolated subset

N: # annotations

β: interpolation factor

CIS centre for intelligent sensing

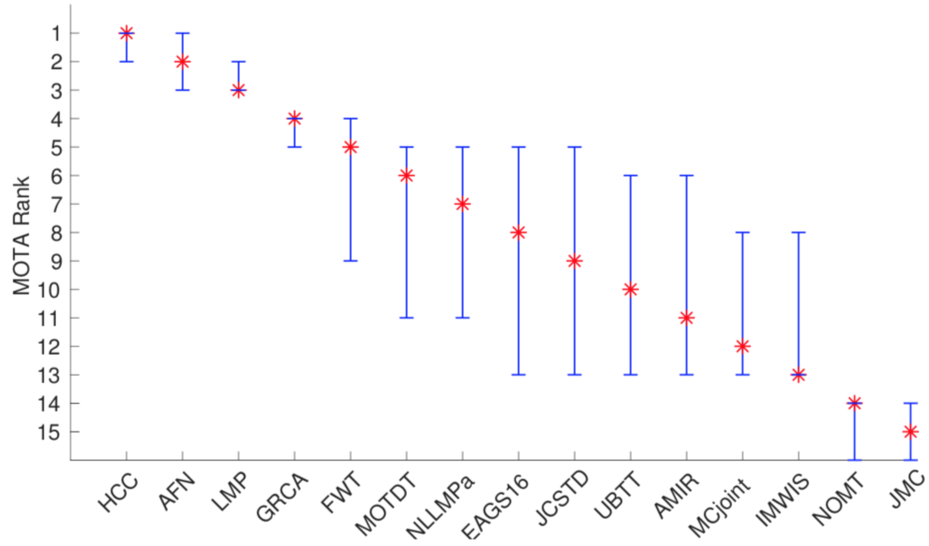
Queen Mary University of London

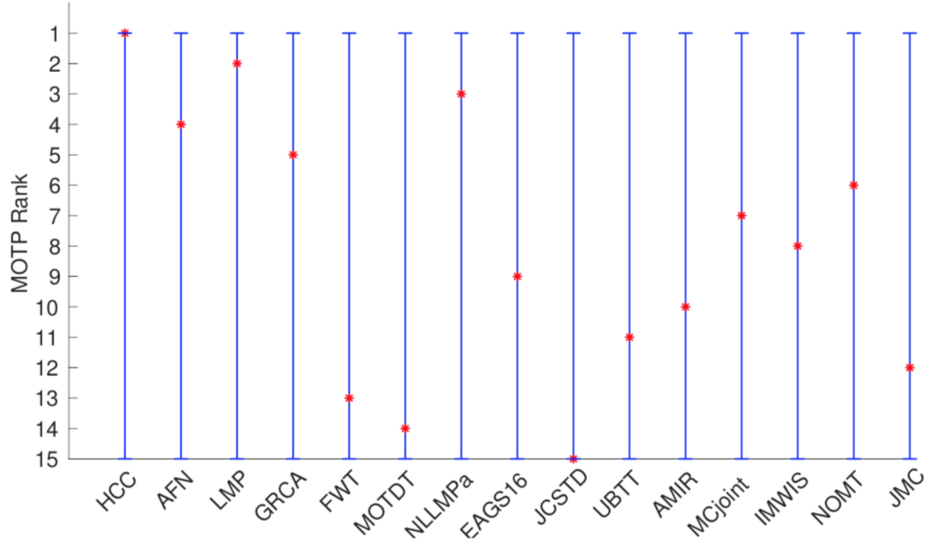
Results on MOTB16

- 39.7% annotations are generated trough linear interpolation
- MOTA confidence: 0.22
- MOTP confidence 3.14

We assume identity switches are negligible

Impact on MOTB16





CIS centre for intelligent sensing

Queen Mary University of London

Conclusion

- Interpolation:
  - Makes possible to annotate large scale datasets
  - Makes possible to annotate not visible objects
  - Introduces inaccuracies
- Confidence interval:
  - Allows to take into account annotation inaccuracies
  - From an already annotated dataset
  - With unknown annotation policy
  - Without the need of doing further annotations
- Ranking of methods should be redesign considering confidence intervals

Q&A