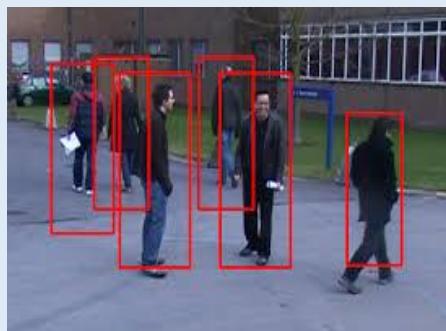


Annual Progress Report on Person Detection, Tracking and Retrieval

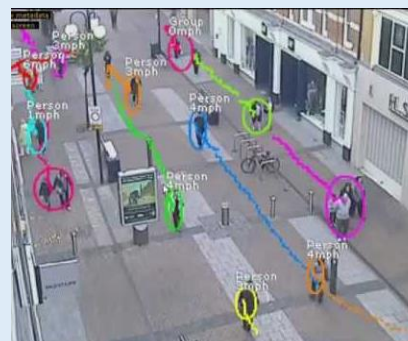
Zhang, Shanshan (张姗姗)

Nanjing University of Science and Technology

Detection



Tracking



Retrieval



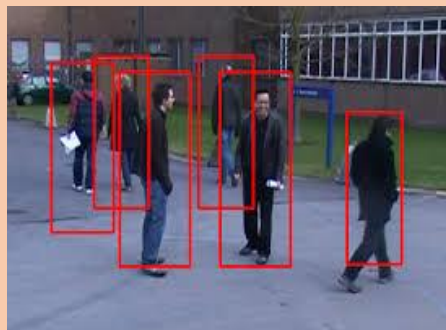


Annual Progress Report on Person Detection, Tracking and Retrieval

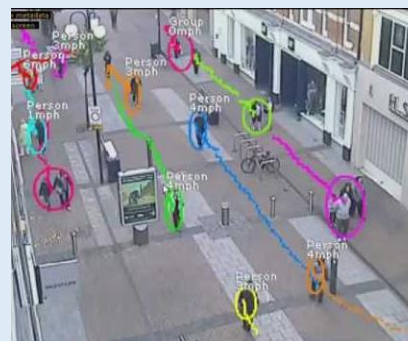
Zhang, Shanshan (张姗姗)

Nanjing University of Science and Technology

Detection



Tracking

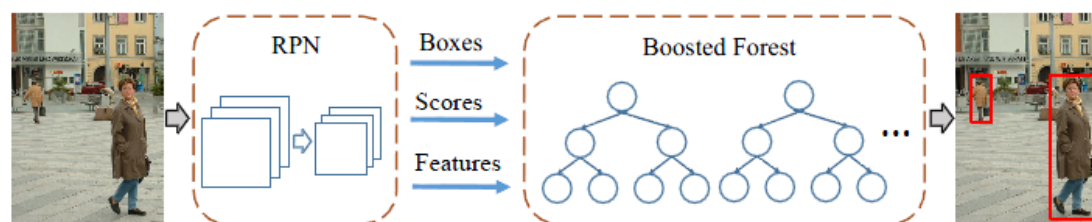


Retrieval

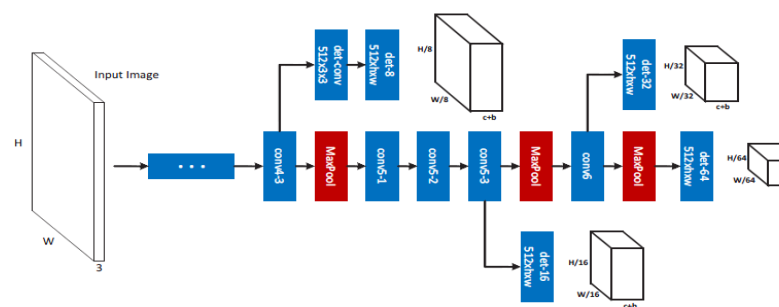
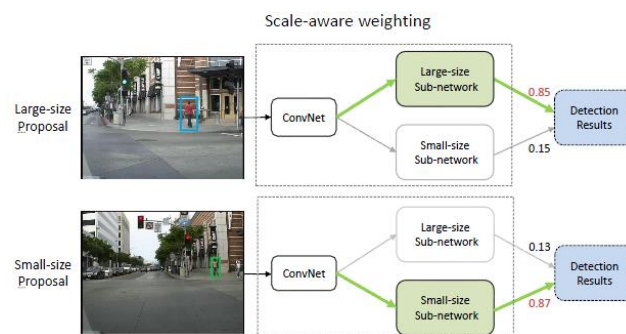


Person detection: CNNs

- CNN + Boosting trees [Zhang et al. CVPR'16, Zhang et al. ECCV'16]



- Multi-scale handling [Li et al. arXiv'16, Cai et al. ECCV'16]



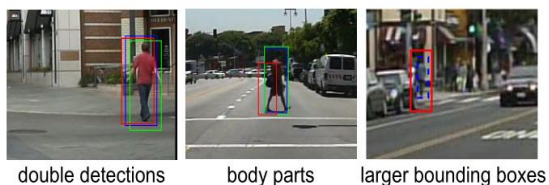
- Semantics [Costea et al. CVPR'16, Du et al. arXiv'16, Zhang et al. CVPR'17]



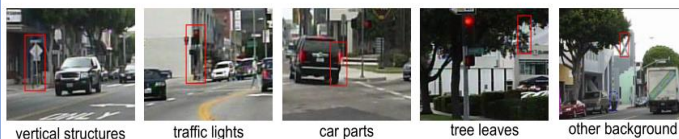
Person detection: Analysis

• Error sources

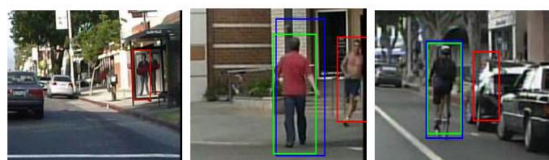
False positives



(a) Localisation errors

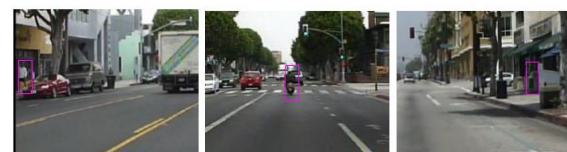


(b) Background errors



(c) Annotation errors

False negatives



• Solutions

- Better aligned training annotations
- CNNs re-scoring

Detector aspect	MR_{-2}^O	MR_{-2}^N
RotatedFilters	19.20	17.22
+ New annotations	16.77	12.96
+ RCNN (VGG)	14.16	10.00

Person detection: CityPersons dataset

- More diverse than previous datasets

	Caltech	KITTI	CityPersons
# country	1	1	3
# city	1	1	18
# season	1	1	3
# person/image	1.4	0.8	7.0
# unique person	1 273	6 336	19 654

- Generalizes better than previous datasets

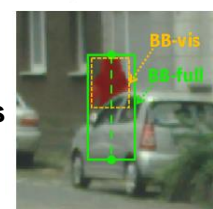
Test	Train	Caltech	KITTI	CityPersons
Caltech		10.27	46.86	21.18
KITTI		10.50	8.37	8.67
CityPersons		46.91	51.21	12.81
INRIA		11.47	27.53	10.44
ETH		57.85	49.00	35.64
Tud-Brussels		42.89	45.28	36.98
mean MR		29.98	38.04	20.95



(a) Image



(b) Segmentation mask

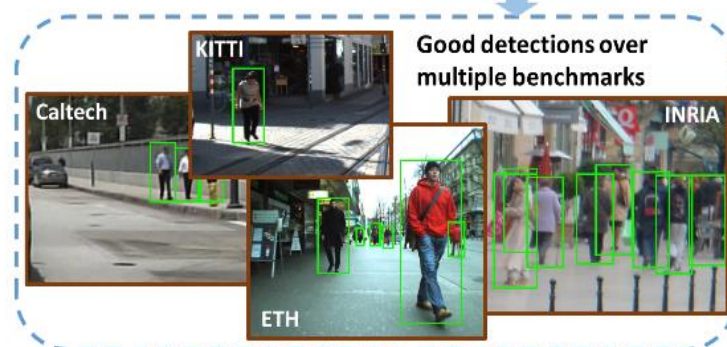


(c) Bounding box anno.

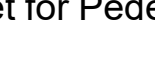
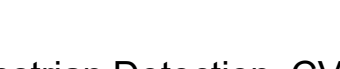


Diverse training data
from CityPersons

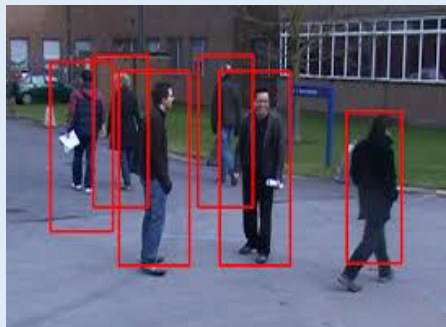
One CNN Model



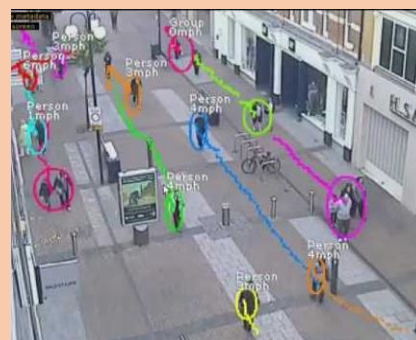
Good detections over
multiple benchmarks



Detection



Tracking

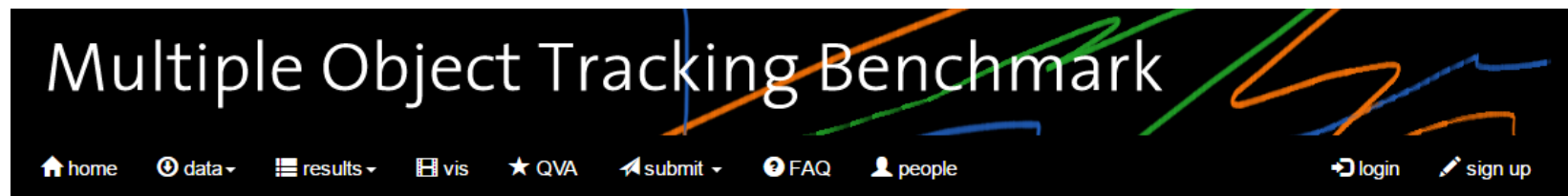


Retrieval



Multi-person tracking: tracking-by-detection

- Multi-cut model is dominating!
 - LMP [Tang et al. CVPR'17]
 - NLLMPa [Levinkov et al. CVPR'17]
 - MCjoint [Keuper et al. arXiv'16]



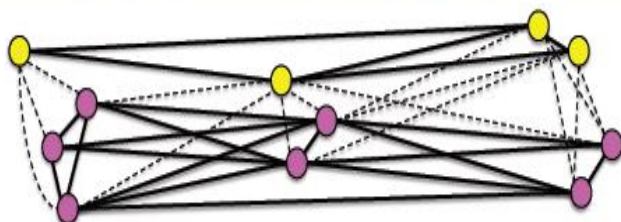
MOT16 Results

Click on a measure to sort the table accordingly. See [below](#) for a more detailed description.

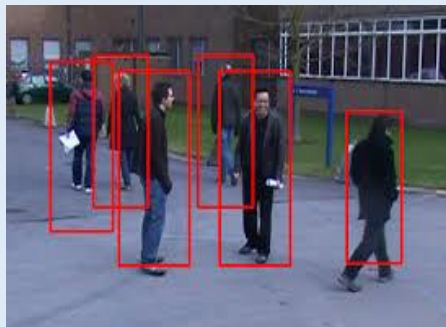
Detections:

Showing only entries that use public detections!

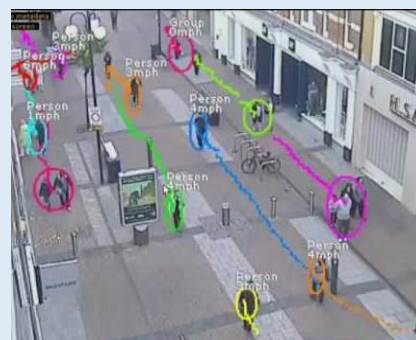
Tracker	Avg Rank	↑MOTA	MOTP	FAF	MT	ML	FP	FN	ID Sw.	Frag	Hz	Detector
HCC 1. <input checked="" type="checkbox"/>	7.3	49.3 ±10.2	79.0	0.9	17.8%	39.9%	5,333	86,795	391 (7.5)	535 (10.2)	0.8	Public
Anonymous submission												
LMP 2. <input checked="" type="checkbox"/>	10.2	48.8 ±9.8	79.0	1.1	18.2%	40.1%	6,654	86,245	481 (9.1)	595 (11.3)	0.5	Public
Anonymous submission												
MLMRF_DL61 3. <input type="checkbox"/> <input checked="" type="checkbox"/>	13.9	48.4 ±9.4	74.3	1.3	18.2%	39.5%	7,849	85,719	491 (9.3)	873 (16.5)	3.0	Public
Anonymous submission												
NLLMPa 4. <input checked="" type="checkbox"/>	9.9	47.6 ±10.6	78.5	1.0	17.0%	40.4%	5,844	89,093	629 (12.3)	768 (15.0)	8.3	Public
E. Levinkov, J. Uhrig, S. Tang, M. Omran, E. Insafutdinov, A. Kirillov, C. Rother, T. Brox, B. Schiele, B. Andres. Joint Graph Decomposition and Node Labeling: Problem, Algorithms, Applications. In CVPR, 2017.												
MDPNN16 5. <input type="checkbox"/> <input checked="" type="checkbox"/>	14.2	47.2 ±7.7	75.8	0.5	14.0%	41.6%	2,681	92,856	774 (15.8)	1,675 (34.1)	1.0	Public
A. Sadeghian, A. Alahi, S. Savarese. Tracking The Untrackable: Learning To Track Multiple Cues with Long-Term Dependencies. In arXiv preprint arXiv:1701.01909, 2017.												
MCjoint 6. <input checked="" type="checkbox"/>	12.8	47.1 ±10.8	76.3	1.1	20.4%	46.9%	6,703	89,368	370 (7.3)	598 (11.7)	0.6	Public
Anonymous submission												



Detection



Tracking



Retrieval



Person retrieval (re-ID)

- Image based



- Video based



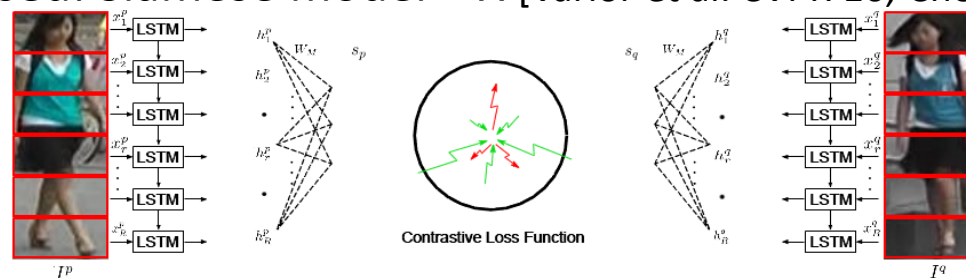
Person retrieval (re-ID)

- Metric learning [Yang et al. AAAI'16, Zhu et al. IJCAI'16, You et al. CVPR'16]

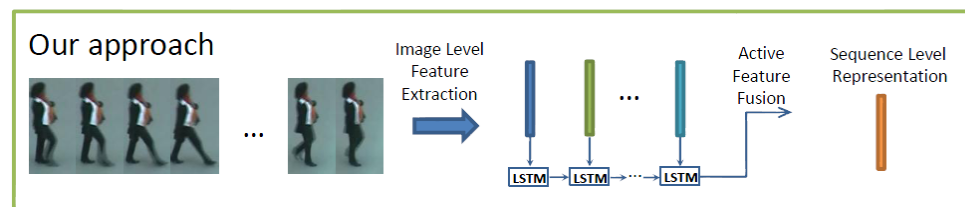
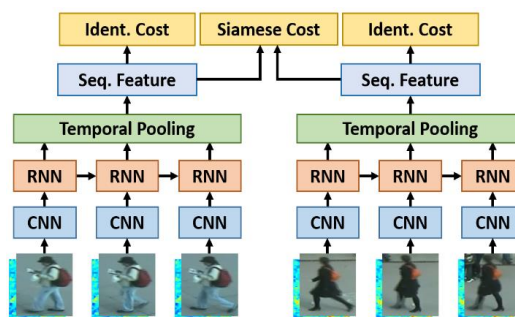
$$d(x_i, x_j) = (x_i - x_j)^T \mathbf{M} (x_i - x_j)$$

- Deep learning

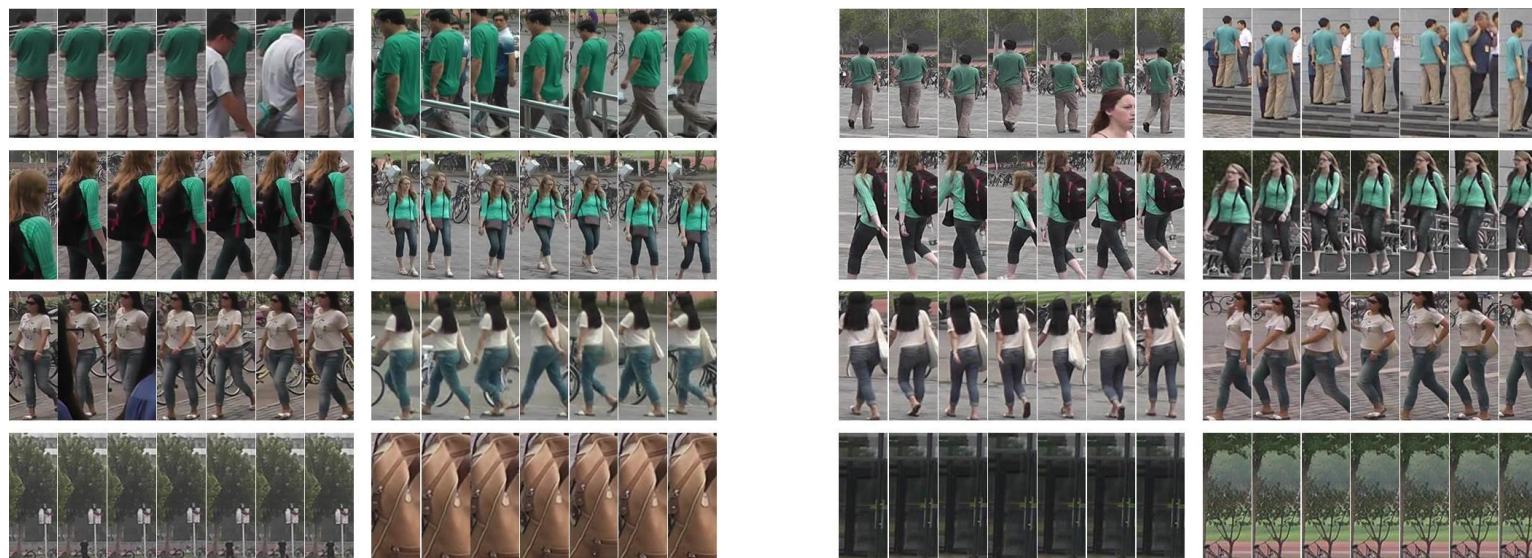
- Image based: Siamese model + X [Varior et al. CVPR'16, Cheng et al. CVPR'16, Su et al. ECCV'16]



- Video based: pooling over frames [McLaughlin et al. CVPR'16, Yan et al. ECCV'16]



Person retrieval (re-ID): MARS dataset

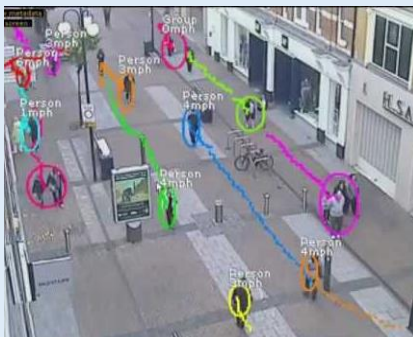


Dataset	MARS	iLIDS	PRID	3DPES	ETH	CUHK03	VIPeR	Market-1501
#ID	1,261	300	200	200	146	1,360	632	1,501
#tracklets	20,478	600	400	1,000	146	-	-	-
#bboxes	1,191,003	43,800	40k	200k	8,580	13,164	1,264	32k
#distractors	3,248	0	0	0	0	0	0	0
#cam./ID	6	2	2	8	1	2	2	6
Produced by	DPM+GMMCP	hand	hand	hand	hand	hand	hand	DPM
Evaluation	mAP+CMC	CMC	CMC	CMC	CMC	CMC	CMC	mAP+R1 precision

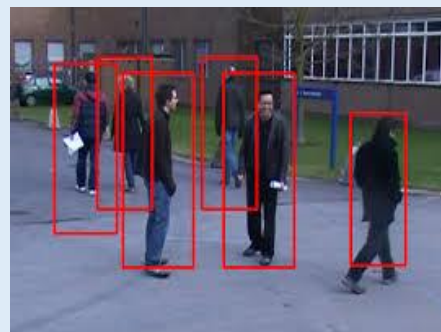
Summary

- We care much about people in image/video data.
- What is pushing the performance?
 - Deep learning
 - Large-scale data
- We are family: detection, tracking and retrieval.

Tracking



Detection



Retrieval





Thank you for your attention!