Lang MA, Qiaoyong ZHONG, Yingying ZHANG, Di XIE, Shiliang PU, 2021. Associative affinity network learning for multi-object tracking. *Frontiers of Information Technology & Electronic Engineering*, 22(9):1194-1206. https://doi.org/10.1631/FITEE.2000272

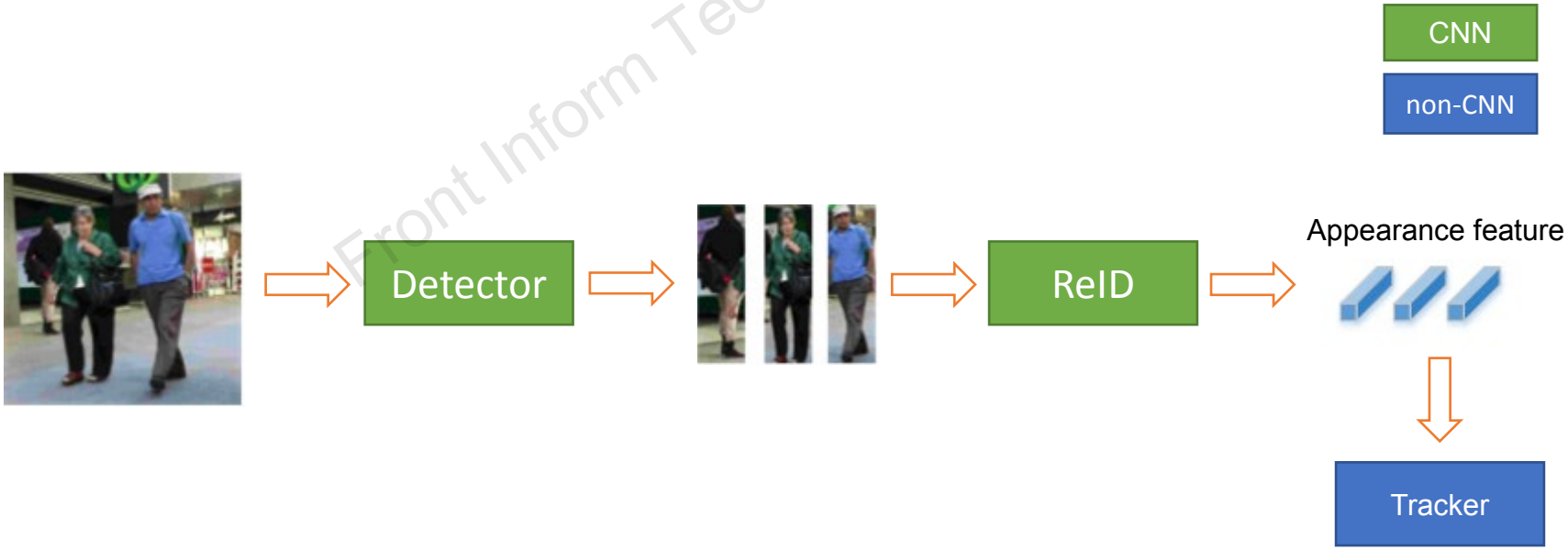
Associative affinity network learning for multi-object tracking

Key words: Multi-object tracking; Deep neural network; Affinity learning

Corresponding author: Liang MA
E-mail: maliang6@hikvision.com
ORCID: https://orcid.org/0000-0003-4228-258X

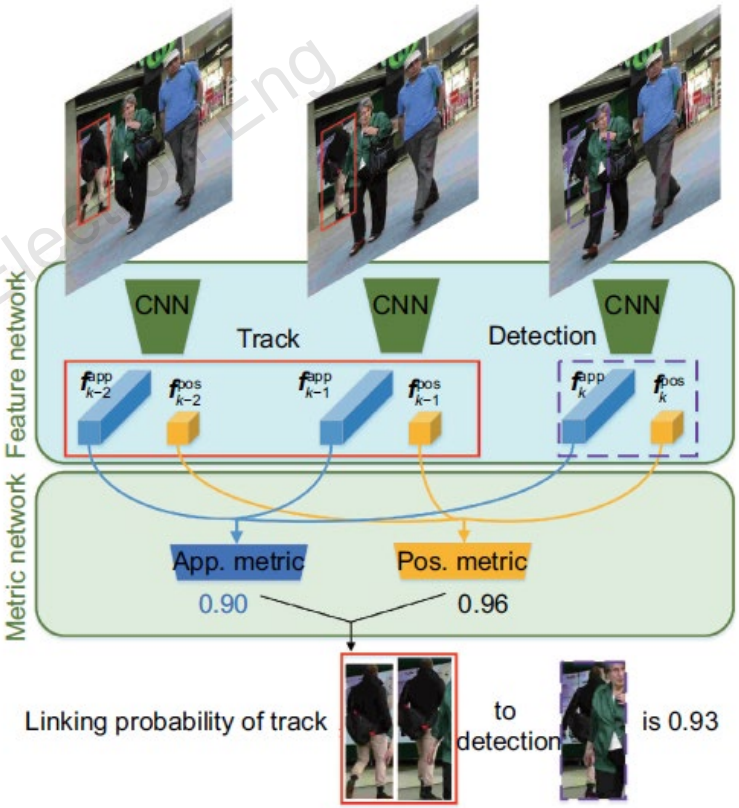
Motivation

- Detector+ReID
  - Less efficiency
  - Not end-to-end trainable
  - Not suitable for flawed detection or targets under occlusion



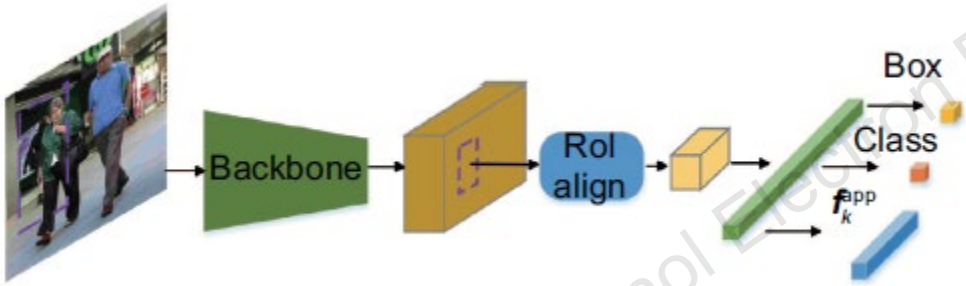
Method

- Associative affinity network
  - End-to-end trainable
  - Directly estimate the associative affinity of a track-detection
  - Perform single-object tracking
  - Alleviate flawed detection and targets under occlusion

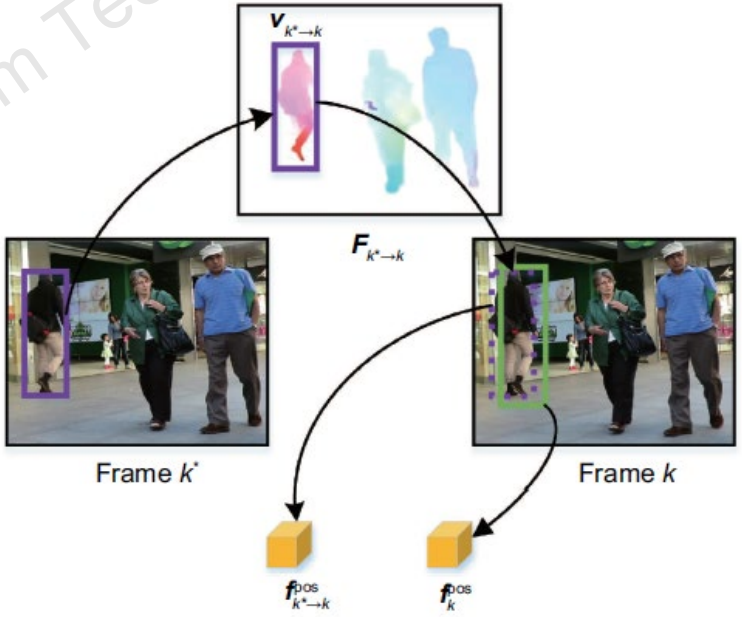


Method: feature network

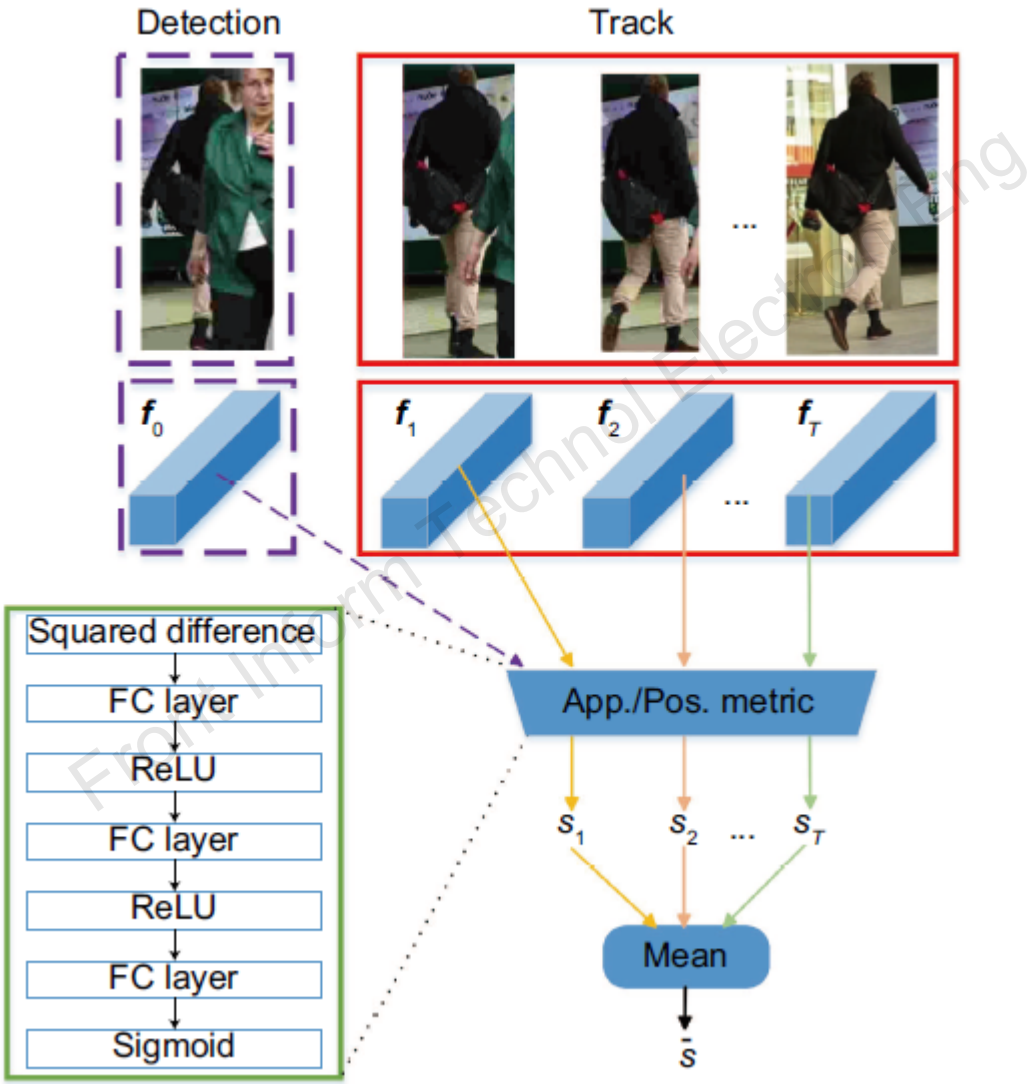
Appearance feature network



Position feature alignment



Method: metric network



MOT results

Table 5 Results on the MOT16 and MOT17 test datasets (trainval, App.+IoU)

Dataset	Tracker	MOTA↑ (%)	IDF1↑ (%)	MT↑	ML↓	FP↓	FN↓	IDS↓	Hz↑
MOT16	UMA (Yin et al., 2020)	50.5	52.8	17.8%	33.7%	7587	81 924	685	5.0
	CRF_TRACK (Xiang J et al., 2021)	50.3	**54.4**	18.3%	35.7%	7148	82 746	702	1.5
	Tracktor++ (Bergmann et al., 2019a)	**54.4**	52.5	19.0%	36.9%	**3280**	79 149	682	1.5
	LMP (Tang et al., 2017)	48.8	51.3	18.2%	40.1%	6654	86 245	595	0.5
	GCRA (Ma et al., 2018)	48.2	48.6	12.9%	41.1%	5104	88 586	821	2.8
	MOTDT (Chen L et al., 2018)	47.6	50.9	15.2%	38.3%	9253	85 431	792	**20.6**
	NOMT (Choi, 2015)	46.4	53.3	18.3%	41.4%	9753	87 565	**359**	2.6
	STAM (Chu Q et al., 2017)	46.0	50.0	14.6%	43.6%	6895	91 117	473	0.2
	MHT_DAM (Kim et al., 2015)	45.8	46.1	16.2%	43.2%	6412	91 758	590	0.8
	LINF1 (Fagot-Bouquet et al., 2016)	41.0	45.7	11.6%	51.3%	7896	99 224	430	4.2
	Ours	53.8	53.4	**21.9%**	**32.0%**	6535	**76 880**	756	6.6
	Tracktor++_v2 (Bergmann et al., 2019b)	56.2	54.9	20.7%	35.8%	**2394**	76 844	**617**	1.6
	Ours (private)	**59.9**	**60.3**	**28.9%**	**22.3%**	11 018	**61 217**	939	5.8
MOT17	UMA (Yin et al., 2020)	53.1	54.4	21.5%	31.8%	22 893	239 534	2251	5.0
	CRF_TRACK (Xiang J et al., 2021)	53.1	53.7	**24.2%**	**30.7%**	27 194	234 991	2518	1.4
	FAMNet (Chu P and Ling, 2019)	52.0	48.7	19.1%	33.4%	14 138	253 616	3072	–
	Tracktor++ (Bergmann et al., 2019a)	53.5	52.3	19.5%	36.6%	**12 201**	248 047	2072	1.5
	DAN (Sun et al., 2021)	52.4	48.7	21.4%	**30.7%**	25 423	234 592	6431	6.3
	TAT (Shen et al., 2018)	51.5	–	20.6%	35.5%	–	–	2593	–
	FWT (Henschel et al., 2018)	51.3	47.6	21.4%	35.2%	24 101	247 921	2648	0.2
	jCC (Keuper et al., 2016)	51.2	**54.5**	20.9%	37.0%	25 937	247 822	**1802**	1.8
	MOTDT (Chen L et al., 2018)	50.9	52.7	17.5%	35.7%	24 069	250 768	2474	**18.3**
	MHT_DAM (Kim et al., 2015)	50.7	47.2	20.8%	36.9%	22 875	252 889	2865	0.9
	Ours	**53.9**	54.3	23.7%	32.0%	27 656	**230 042**	2386	5.4
	Tracktor++_v2 (Bergmann et al., 2019b)	56.3	55.1	21.1%	35.3%	**8866**	235 449	**1987**	1.5
	Ours (private)	**58.8**	**59.5**	**27.8%**	**24.5%**	30 888	**198 864**	2880	5.8
	Ours (private) w/o optical flow	58.6	59.2	**27.8%**	**24.5%**	31 467	199 275	3024	**9.8**

↑: a higher score means a better result; ↓: a lower score means a better result. The best results are in bold

Summary

- An associative affinity network (AAN) for the task of multi-object tracking has been proposed. Appearance and position features and their affinity metrics have been learned jointly in an end-to-end manner.

- AAN can perform single-object tracking to maintain trajectory consistency in the presence of missed detections.

- A simple multi-object tracker with AAN achieved competitive performance on the MOT16 and MOT17 datasets.