Visio-Temporal Attention for Multi-Camera Multi-Target Association

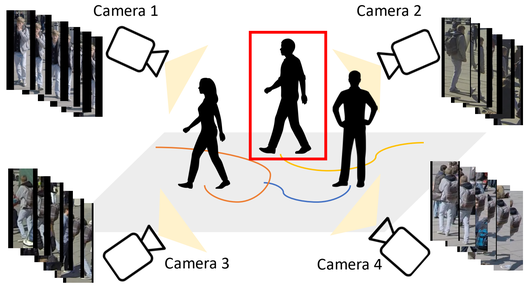
Yu-Jhe Li Xinshuo Weng Yan Xu Kris Kitani

Carnegie Mellon University — The Robotics Institute

2021 ICCV OCTOBER 11-17 VIRTUAL

Goal

Identify the person with multiple synchronized cameras (views).

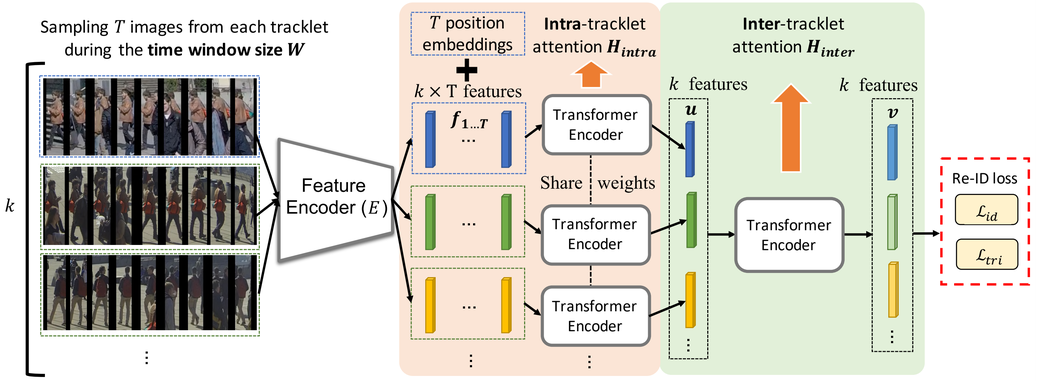


Challenges

Besides occlusion, the action type spans walking, standing, squating, kneeling, and etc.

Methodology

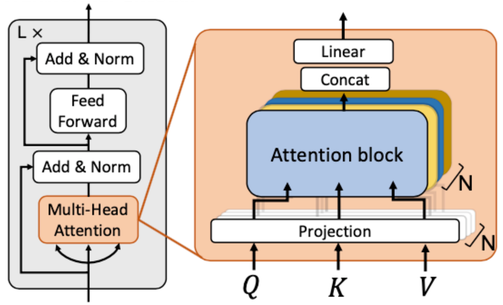
Video-based Re-ID and MTMC tracking with transformer encoders



- **Intra-tracklet attention :**
  To learn a person-specific motion and appearance feature with a transformer encoder plus positional encodings.

- **Inter-tracklet attention :**
  To compute a discriminative feature representation by taking into account all other time synchronized tracklets across all camera views with a transformer encoder.

Transformer Encoder



- **Full objective:**
  Re-ID loss (classification and triplet losses) is used to update the entire network in the end-to-end manner.

Experiments

- Experimental datasets: ***ConstructSite***, ***WILDTRACK***
- Quantitative results on Re-ID (*indicates no source code):

Method	Source	ConstructSite			WILDTRACK		
		Synchronized & overlapping					
		Rank1	Rank5	mAP	Rank1	Rank5	mAP
ResNet-50 [18]	CVPR16	69.7	94.4	73.1	70.4	88.9	57.5
ETAP-Net [61]	CVPR18	72.3	93.4	71.3	71.2	88.7	58.4
STA [13]*	AAAI19	-	-	-	-	-	-
GLTR [21]	ICCV19	78.2	94.6	75.1	75.8	89.2	59.7
TKP [17]	ICCV19	77.2	94.0	73.8	77.6	91.3	59.6
COSAM [44]	ICCV19	76.3	94.4	73.1	77.5	91.2	59.3
NVAN [28]	BMVC19	85.0	95.4	78.0	80.4	92.6	66.3
VKD [33]	ECCV20	85.6	96.0	80.1	79.9	92.5	66.1
AP3D [16]	ECCV20	85.4	95.8	80.5	80.3	92.1	67.0
Ours ($L = 1$)	default	94.2	99.1	90.8	85.1	96.5	71.6
Ours ($L = 3$)	default	**94.7**	**99.3**	91.0	**85.5**	96.8	**72.0**
Ours ($L = 5$)	default	94.5	99.2	**91.1**	85.4	**96.9**	71.7

- Quantitative results on MTMC tracking:

Method	ConstructSite		
	IDF1	IDP	IDR
GT tracklets+ ResNet-50 [18]	66.50	65.42	66.71
GT tracklets+ NVAN [28]	84.72	87.15	82.63
GT tracklets+ VKD [33]	85.20	84.74	86.91
GT tracklets+ AP3D [16]	84.48	83.64	85.34
GT tracklets+ Ours	**92.38**	**91.31**	**93.47**
DeepSort [59] + ResNet-50 [18]	30.05	21.84	40.16
DeepSort [59] + NVAN [28]	49.16	40.01	56.58
DeepSort [59] + VKD [33]	47.35	36.48	51.31
DeepSort [59] + AP3D [16]	47.56	38.04	53.50
DeepSort [59] + Ours	**62.69**	**61.97**	**63.44**