AAAI-19:
Thirty-Third AAAI Conference on Artificial Intelligence
January 27 – February 1, 2019, Hilton Hawaiian Village, Honolulu, Hawaii, USA

Multi-scale 3D Convolution Network for Video Based Person Re-Identification

Jianing Li, Shiliang Zhang , Tiejun huang

PEKING UNIVERSITY 北京大学

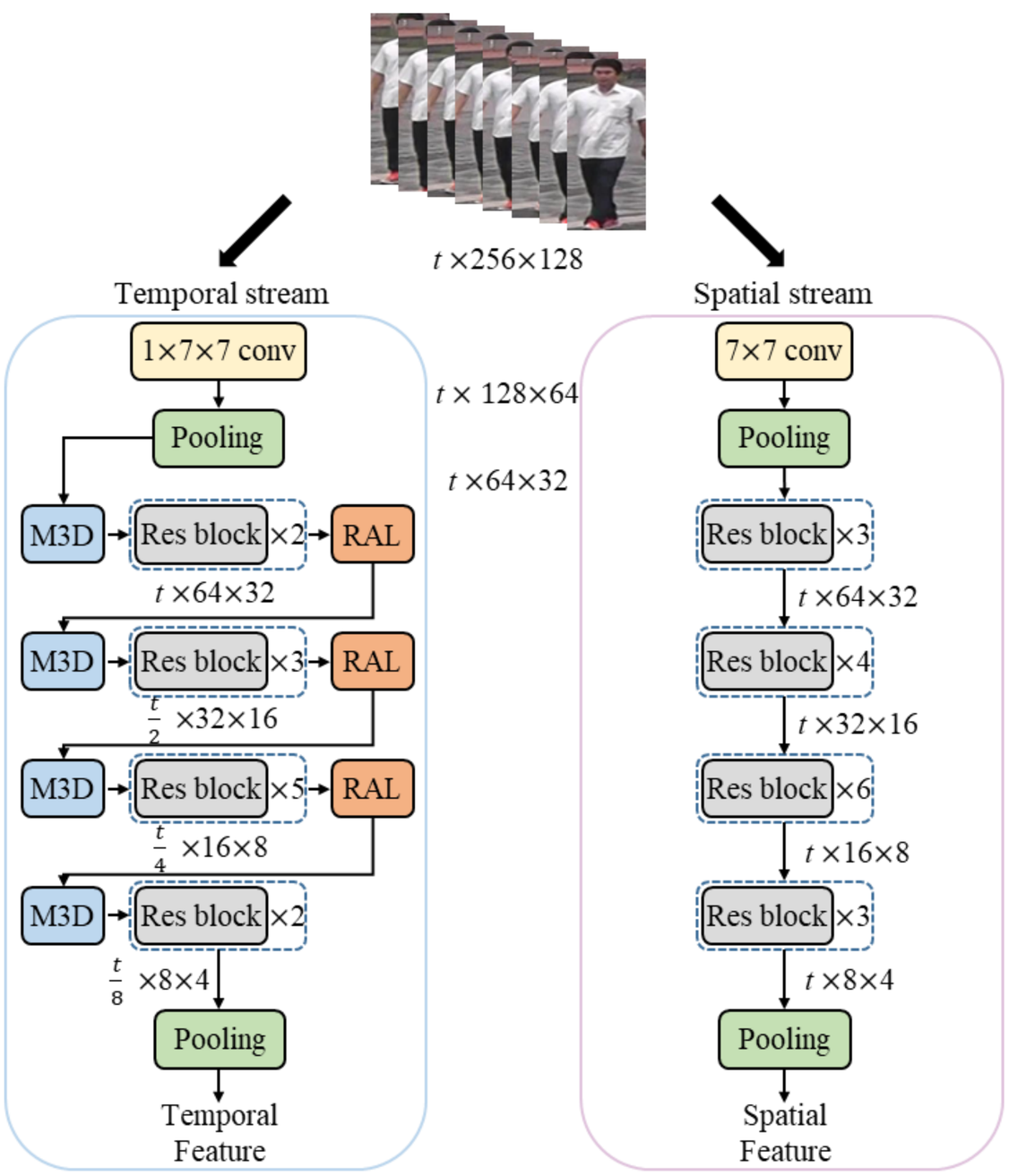
Motivation

- Temporal cues are important for video ReID
- Existing 3D CNNs have small receptive field and too many parameters
- Low quality frame is unavoidable in real scene

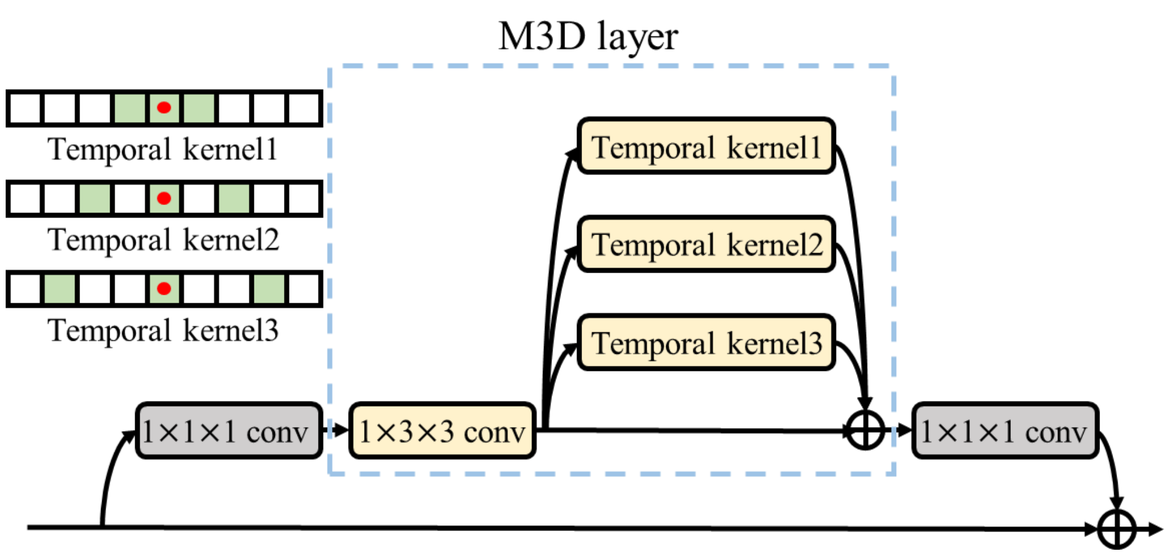
Contribution

- Propose M3D Convolution model to learn multi-scale temporal cues
- Propose RAL to refine learned temporal feature
- Introduce a two-stream architecture to learn complementary spatial temporal representation

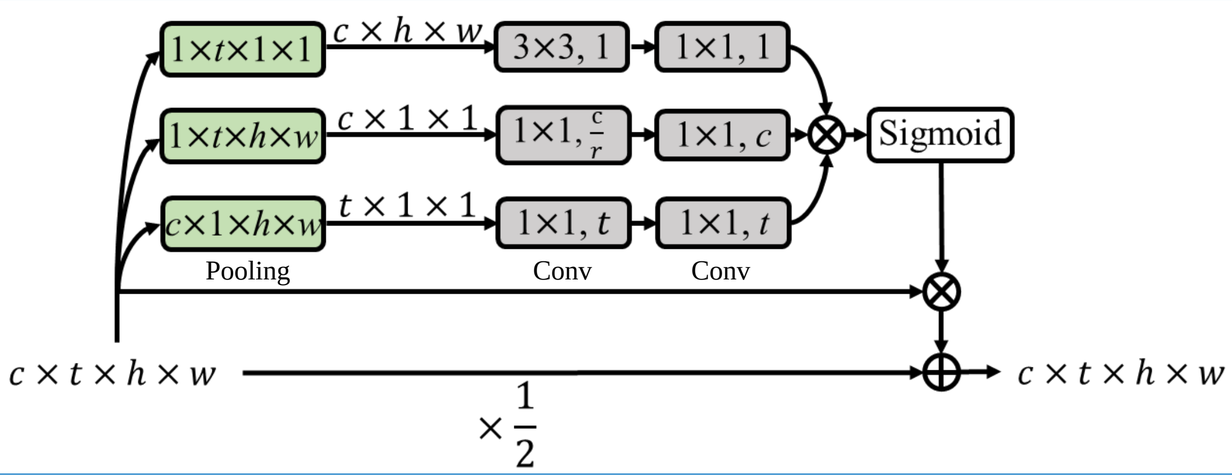
Two-Stream M3D network



Multi-scale 3D Convolution



Residual Attention Layer



Experiment

- Comparison with 3D convolution methods

Method	Input Frames	mAP	r1	Speed	Params
2D CNN	1	62.54	76.43	796 frame/s	95.7MB
I3D	8	62.84	76.62	81.0 clip/s	186.3MB
	16	61.58	75.11	38.7 clip/s	
P3D-A	8	60.69	75.08	90.1 clip/s	110.9MB
	16	60.52	75.69	46.9 clip/s	
P3D-B	8	67.03	79.06	93.9 clip/s	110.9MB
	16	65.07	77.63	48.7 clip/s	
P3D-C	8	67.06	79.08	87.6 clip/s	110.9MB
	16	65.17	79.44	45.4 clip/s	
M3D	8	**69.90**	**81.01**	**98.3 clip/s**	**99.9MB**
	16	66.23	80.13	49.1 clip/s	

- Ablation study

Dataset	*MARS*		*PRID*	*iLIDS-VID*
Method	mAP	r1	r1	r1
2D baseline	62.54	76.43	82.02	49.33
M3D	69.90	81.01	87.64	70.00
M3D+RAL(s)	71.04	82.19	89.89	71.33
M3D+RAL(t)	70.66	81.81	88.76	71.33
M3D+RAL(c)	71.30	82.13	89.89	72.00
M3D+RAL	71.76	82.79	91.03	72.67
Two-stream M3D	**74.06**	**84.39**	**94.40**	**74.00**

- Comparison on MARS

Method	mAP	r1	r5	r20
DCF (Li et al. 2017a)	56.05	71.77	86.57	93.08
SeeForest (Zhou et al. 2017)	50.70	70.60	90.00	97.60
DRSA (Li et al. 2018)	65.80	82.30	-	-
DuATM (Si et al. 2018)	67.73	81.16	92.47	-
LSTM (Yan et al. 2016)	61.58	76.11	85.30	92.68
A&O (Simonyan et al. 2014)	63.39	77.11	88.41	94.60
Two-stream M3D	**74.06**	**84.39**	**93.84**	**97.74**

- Comparison on PRID&iLIDS-VID

Dataset	*PRID*		*iLIDS-VID*	
Method	r1	r5	r1	r5
IDE+XQDA (Zheng et al. 2016)	77.30	93.50	53.00	81.40
SeeForest (Zhou et al. 2017)	79.40	94.40	55.20	86.50
AMOC (Liu et al. 2017a)	83.70	98.30	68.70	94.30
QAN (Liu et al. 2017b)	90.30	98.20	68.00	86.80
DRSA (Li et al. 2018)	93.20	-	**80.20**	-
Two-stream M3D	**94.40**	**100.00**	74.00	**94.33**

Contact Information

Jianing Li(ljn-vmc@pku.edu.cn)
Shiliang Zhang(slzhang.jdl@pku.edu.cn)
The source code have been released.