Human Centric Spatio-Temporal Action Localization

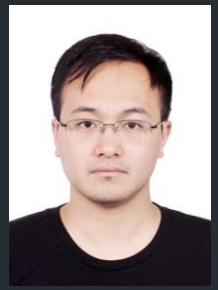
Megvii(Face++) Team

Member

Jianwen Jiang

Yu Cao

Lin Song



Shiwei Zhang

Yunkai Li

Ziyao Xu

Qian Wu

Chuang Gan

Chi Zhang

Gang Yu

AVA

80 atomic visual actions: 17 person to person, 49 person to object, 14 pose

299 movies: 235 training and 64 val test videos,131 for testing

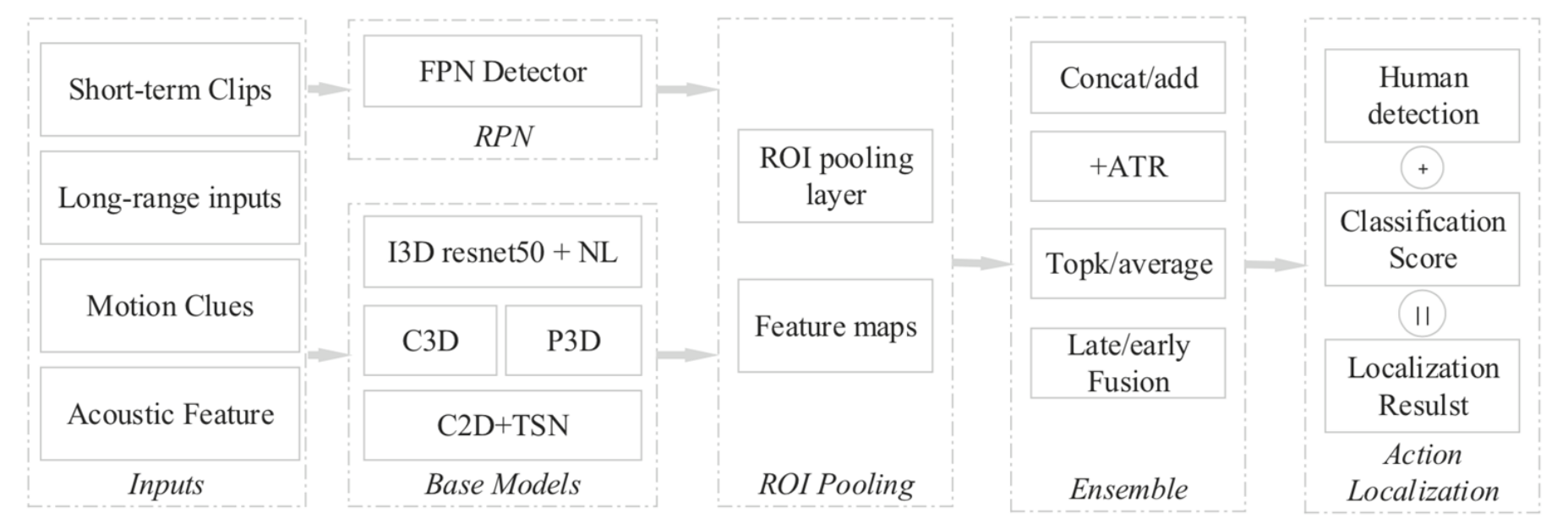
15th min to 30th min (length: 15mins)

Frame-MAP on 60 classes

Method (Multi-clues)

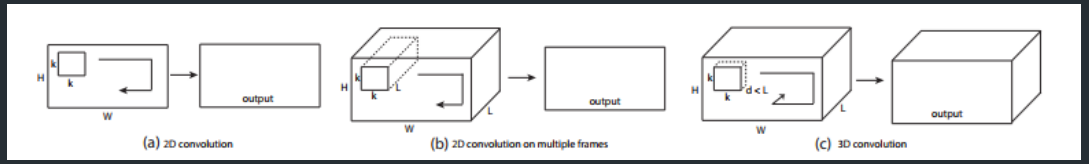
- RGB
- Optical flow
- Acoustic features

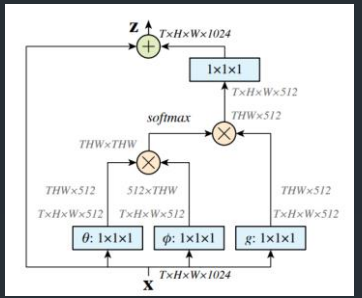
Method (Multi-clues)

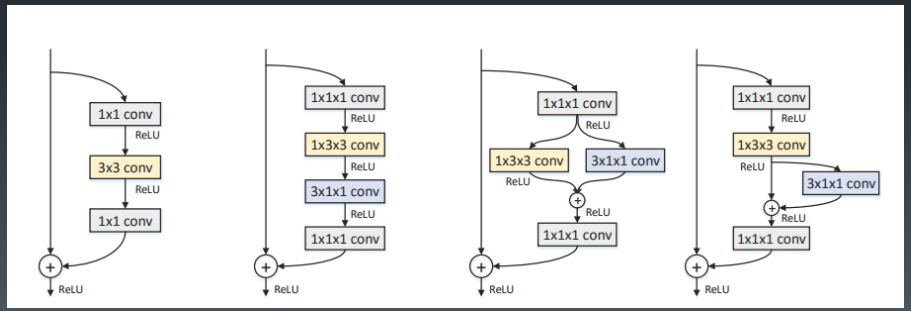


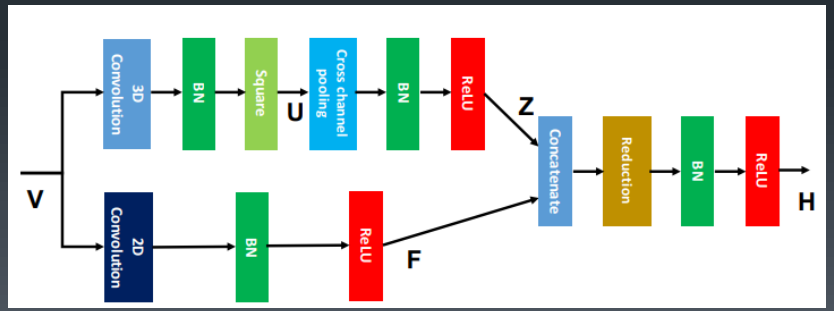
Short-term clips

- C3D
- P3D
- Appearance-and-Relation Networks
- Non-local





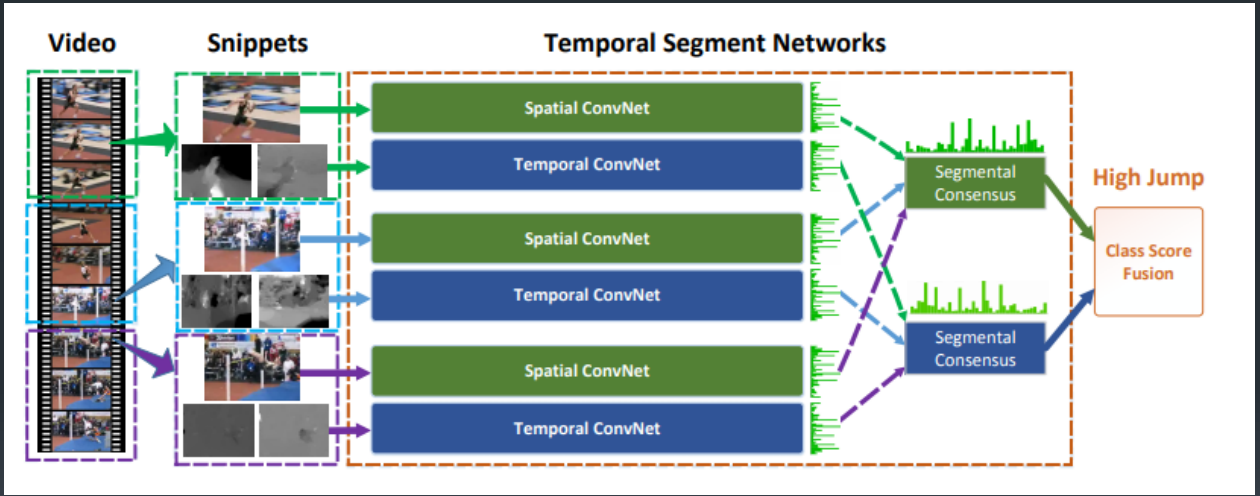




[1] D. Tran, L. Bourdev, R. Fergus, L. Torresani, and M. Paluri. Learning spatiotemporal features with 3d convolutional networks. In ICCV, pages 4489–4497. IEEE, 2015. [2] . Qiu, T. Yao, and T. Mei. Learning spatio-temporal representation with pseudo-3d residual networks. In ICCV, 2017. [3] L. Wang, W. Li, W. Li, and L. Van Gool. Appearance-and-relation networks for video classification. CVPR, 2018. [4] X. Wang, R. Girshick, A. Gupta, and K. He. Non-local neural networks. arXiv preprint arXiv:1711.07971, 2017.

Long-term clips

- 2D+TSN



L. Wang, Y. Xiong, Z. Wang, Y. Qiao, D. Lin, X. Tang, and L. V. Gool. Temporal segment networks: Towards good practices for deep action recognition. ECCV, 22(1):20–36, 2016.

Actor-Target Relationship

Motivation

- 49 actor-target relationship classes
- ROI pooling focuses on human feature
- Background feature not utilized

Ensemble

Model	Input	Modality	Operatoin	mAP
I3D Resnet50 + NL	3,20,224,224	RGB	-	19.33
	3,20,224,224	RGB	ATR	20.01
	3,20,224,224	Flow	-	16.64
	3,20,360,400	RGB	-	19.86
	3,20+20,224,224	RGB+Flow	add	21.66
P3D199	3,20+20,224,224	RGB+Flow	concat	17.87
Resnet152	3,20,224,224	RGB	TSN	14.68
VGG16	5,96,64	Audio	-	6.8
I3D Resnet101 + NL	3,20,224,224	Flow	-	18.10
P3D199	3,20+20,224,224	Flow	-	15.17
Resnet50 2D	3,16,224,224	Flow	AVG	12.48

Computer Vision Only	25.63
Full Track (w audio)	25.75

Thank you!

yugang@megvii.com