TAIWAN TECH
NATIONAL TAIWAN UNIVERSITY OF SCIENCE AND TECHNOLOGY

MULTIPLE PATH SEARCH FOR ACTION TUBE DETECTION IN VIDEOS

Erick Hendra Putra Alwando and Yie-Tarng Chen and Wen-Hsien Fang

Problem

We address the problem of localizing and classifying actions in video.

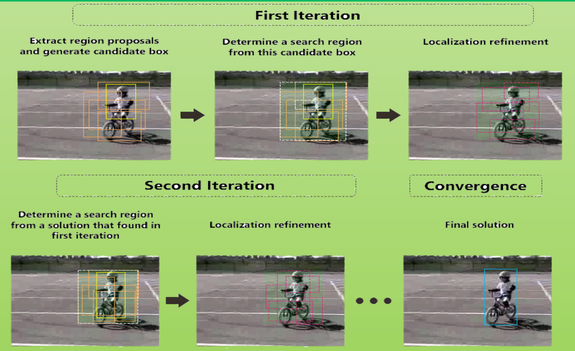
- Action path generation using dynamic programming approach still **inefficient**.
- **Inaccurate** object localization in the existing CNN-based methods.

Contributions

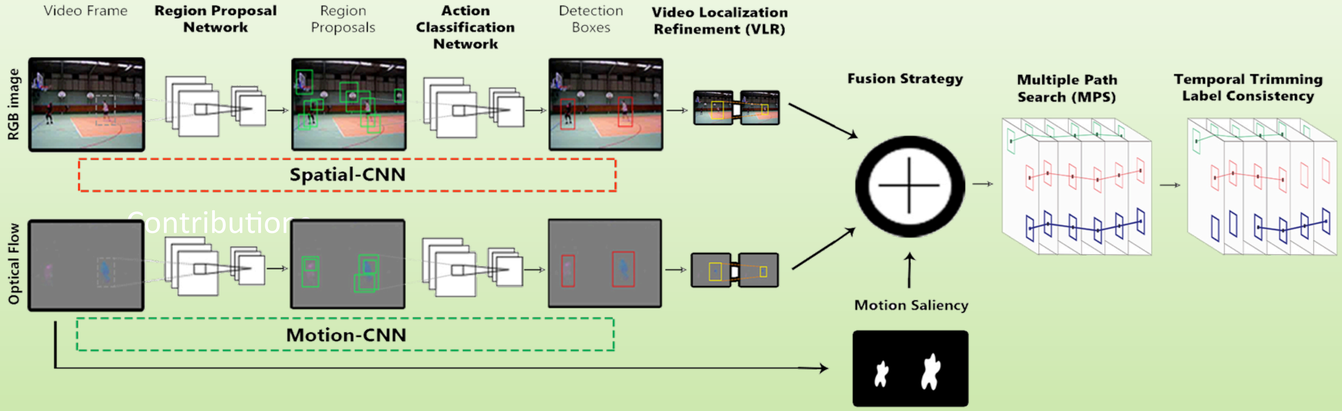
1. A novel **low-cost dynamic programming -like algorithm** for efficient action path generation.
2. A new **fusion strategy** for reducing camera motion.
3. An **iterative video localization refinement scheme** for refine inaccurate bounding boxes.

Video Localization Refinement

- Simple
- Iterative
- Learning-free
- CNN-compatible
- Efficient
- Less parameters

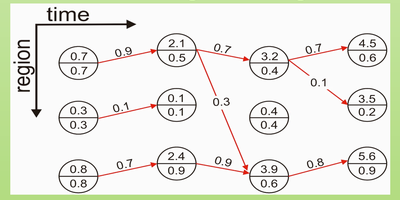


System Diagram



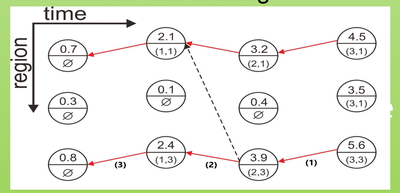
Multiple Path Search

Forward-Message-Passing



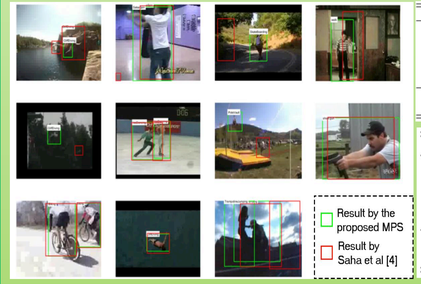
Store K largest paths instead only one with the maximum score.

Backward-Path-Tracing



Trace back based on the accumulated score, in this step we have more option to chose a path.

Experimental Result



Method	0.05	0.1	0.2	0.3	0.4
Mettes *et al* [14]	-	-	32.40	-	-
FAP [7]	42.80	-	-	-	-
Gemert *et al* [12]	58.00	-	37.80	-	-
STMH [6]	54.28	51.68	46.77	37.82	-
Saha *et al* [4]	79.12	76.57	66.75	55.46	46.35
MPS	81.73	78.61	67.96	57.81	47.90
MPS + VLR	**84.62**	**81.39**	**70.36**	**58.44**	**48.86**

UCF-101 dataset

Method	0.1	0.2	0.3	0.4	0.5
ActionTube [5]	-	-	-	-	53.30
Wang *et al* [15]	-	-	-	-	56.40
STMH [6]	-	63.10	63.50	62.20	60.70
Saha *et al* [4]	72.65	72.63	72.59	72.24	71.50
MPS	73.36	73.36	73.27	72.89	72.29
MPS + VLR	**78.97**	**78.93**	**78.85**	**78.63**	**77.35**

J-HMDB dataset

[4] S. Saha, G. Singh, M. Sapienza, P. H. S. Torr, and F. Cuzzolin, "Deep learning for detecting multiple space-time action tubes in videos," in Proceedings of the British Machine Vision Conference, 2016.
[5] G. Gkioxari and J. Malik, "Finding action tubes," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015,pp. 759–768.
[6] P. Weinzaepfel, Z. Harchaoui, and C. Schmid, "Learning to track for spatio-temporal action localization," in Proceedings of the IEEE International Conference on Computer Vision, 2015, pp. 3164–3172.
[7] G. Yu and J. Yuan, "Fast action proposals for human action detection and search," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 1302–1311.