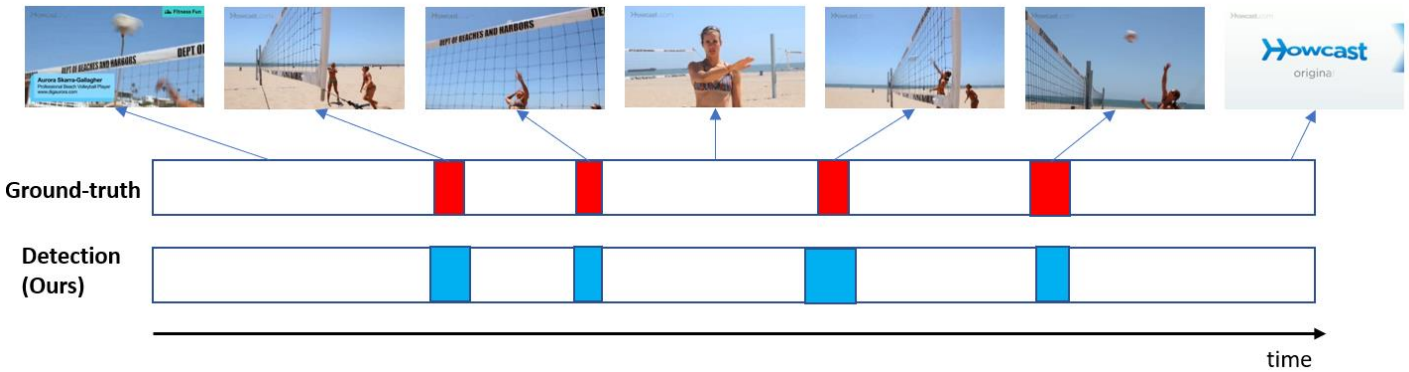
Weakly Supervised Temporal Action Localization with Proxy Metric Modeling

Hongsheng XU, Zihan CHEN, Yu ZHANG , Xin GENG, Siya MI, Zhihong YANG

Frontiers of Computer Science, DOI: 10.1007/s11704-022-1154-1

Problems & Ideas

- Problems of weakly-supervised temporal action localization approaches:
  - Some segments do not possess sufficient action features but still participate in clustering, which will reduce the accuracy of model training.
  - Most methods can not employ the representative vectors of action classes that can identify the relevance of each segment to improve the localization accuracy
- Ideas: A novel proxy metric model to find proxy vectors based on the similarity relationships between video segments. This paper propose a proxy-based metric to cluster the same actions together and separate actions from backgrounds.



The qualitative results of VolleyballSpikingGolfSwing action on THUMOS14.Ground-truth and the localization results of our method are shown in red and blue.

Main Contributions

- Contributions:
  - We apply the proxy losses in weakly-supervised action localization for the first time;
  - Our pseudo segments mechanism only consider the critical part of a video, which are effective for action localization;
  - We propose a novel similarity matrix based on the action proxy map.
  - The proposed method is systematically evaluated on benchmark datasets, which shows its efficacy compared with existing methods.

Supervision	Method	mAP@IoU					AVG
		0.3	0.4	0.5	0.6	0.7	
Full	S-CNN [19]	36.3	28.7	19.0	10.3	5.3	19.9
Full	CDC [20]	40.1	29.4	23.3	13.1	7.9	22.8
Full	R-C3D [21]	44.8	35.6	28.9	-	-	-
Full	TAL-Net [22]	53.2	48.5	42.8	33.8	20.8	39.8
Weak	STPN(UNT) [23]	31.1	23.5	16.2	9.8	5.1	17.1 •
Weak	STPN(I3D) [23]	35.5	25.8	16.9	9.9	4.3	18.5 •
Weak	Liu et al. [25]	41.2	32.1	23.1	15.0	7.0	23.7 •
Weak	W-TALC [24]	40.1	31.1	22.8	-	7.6	25.4 •
Weak	RPN [17]	48.2	37.2	27.9	16.7	8.1	27.6 •
Weak	BMU [12]	46.9	39.2	30.7	20.8	12.5	30.0 ○
Weak	Ours	46.8	39.1	**30.9**	**21.0**	**12.6**	**30.1**

Action localization performance compared on the THUMOS14 dataset. The last column AVG indicates the average mAP at IoU thresholds 0.3:0.1:0.7.