Adversarial Seeded Sequence Growing for Weakly-Supervised Temporal Action Localization

ACMMULTIMEDIA 2019
21 - 25 October 2019, Nice, France

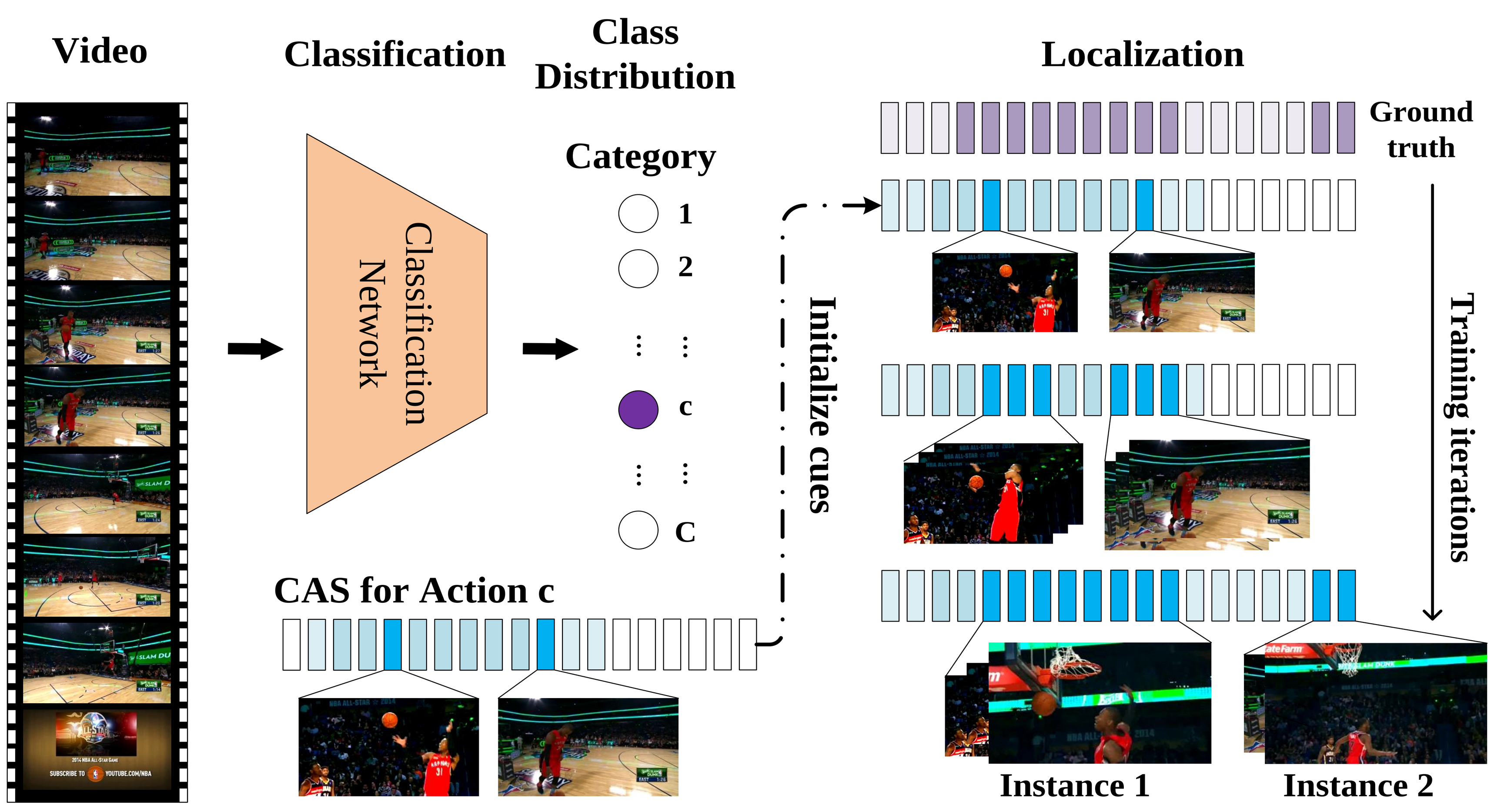
Chengwei Zhang; Futai Zou   Yunlu Xu; Zhanzhan Cheng [23]; Yi Niu; Shiliang Pu   Wu Fei
[1] Shanghai Jiao Tong University   [2] Hikvision Research Institute   [3] Zhejiang University

Program ID: P2B-12   Contact Mail: xuyunlu@hikvision.com

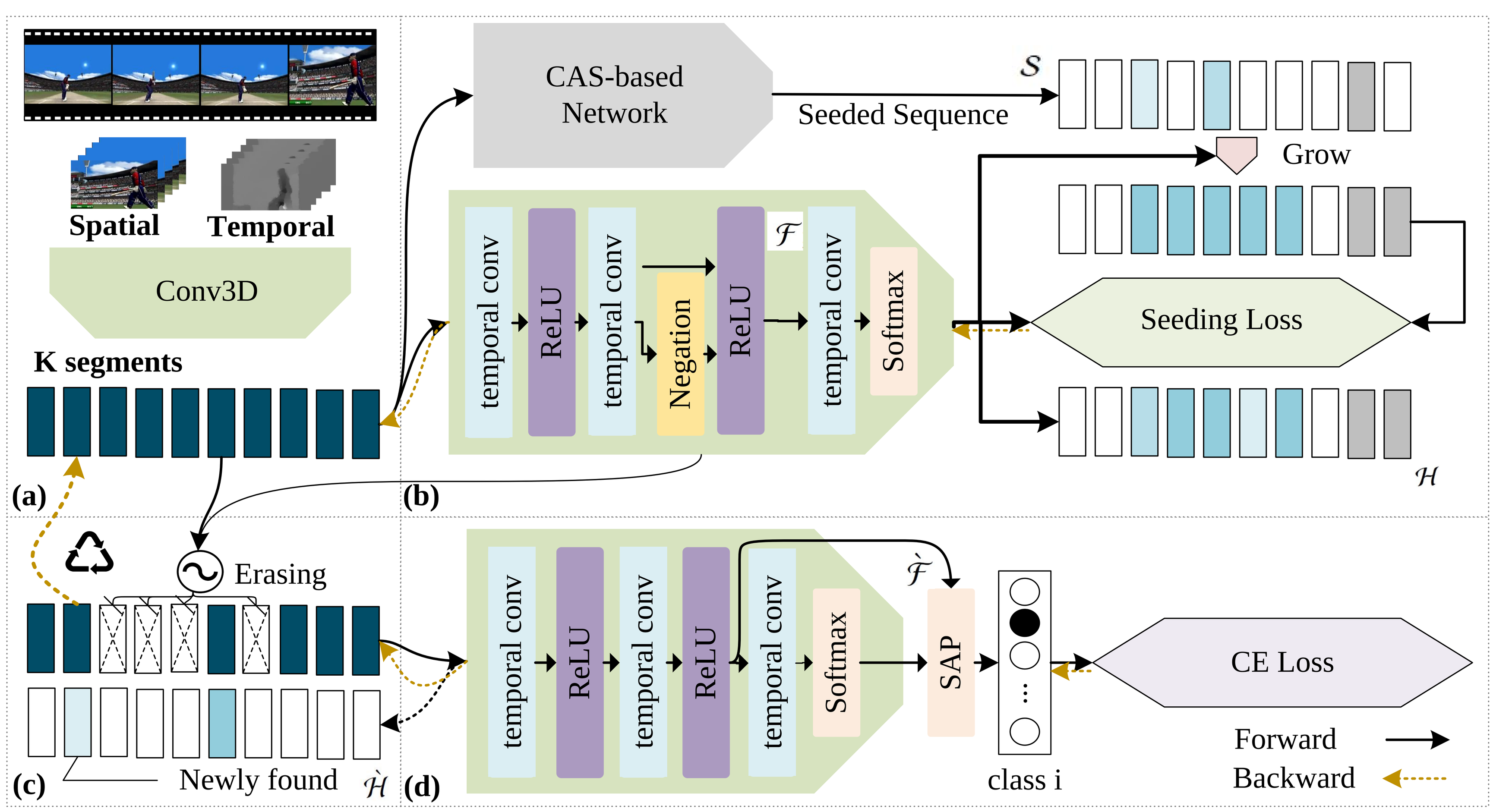
MOTIVATION

- Existing methods on weakly-supervised temporal action recognition mainly work based on **Class Activation Sequence (CAS)**.

- The CAS-based action detector usually localizes actions in untrimmed videos at the **most discriminative** action interval, which often appears in action response peak.



METHOD



- We propose a new weakly-supervised action detection framework called **Adversarial Seeded Sequence Growing (ASSG)** by adversarial learning of a **Seed Sequence Growing (SSG)** network and a **self-adaptive action classification** network.

- We train two modules in an **adversarial** manner, which not only helps grow action occurring durations and also mines trivial or indiscriminative actions.

- Our method achieves impressive performance on the challenging benchmarks, especially on the evaluation of **high IoUs** which is more valuable than that in low IoUs.

$$\mathcal{G}(\mathcal{H}_{c,t},\mathcal{S}_c,\theta_g^c)=\begin{cases} True, & l\in\mathcal{N}(\mathcal{S}_c),\ \mathcal{H}_{c,t}\geq\theta_g^c \\ & \text{and } c=\arg\max_{c'}\mathcal{H}_{c',t} \\ False, & otherwise. \end{cases}$$

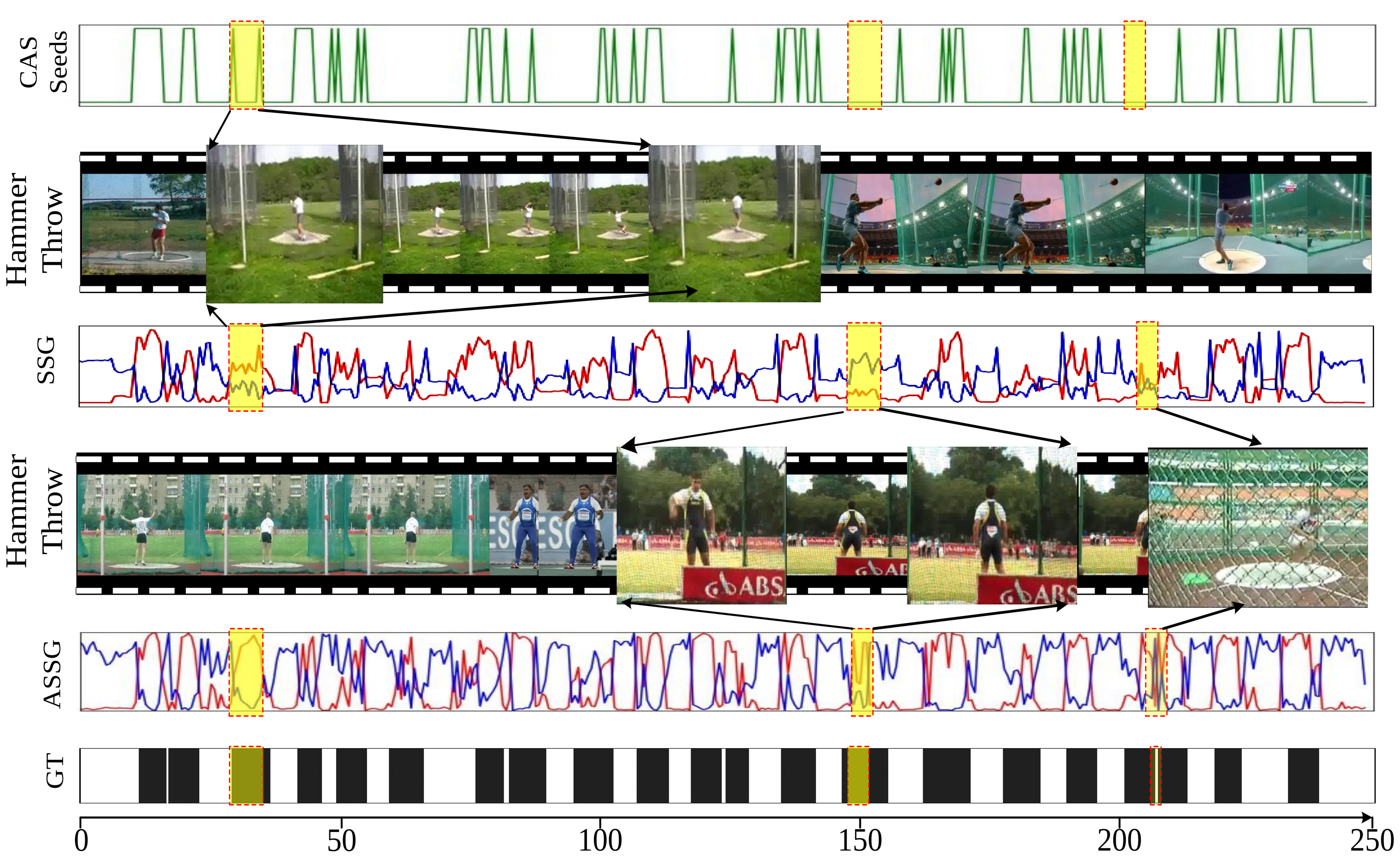
$$SAP(X)=\sum_{t=1}^{N}A_{*,t}(X)\cdot\grave{\mathcal{H}}_{*,t}$$

$$A_{c,t}(X)=\frac{e^{(\sum_{i=1}^{|f_{c,t}(X)|}f_{c,t}^i(X))}}{\sum_{t=1}^{N}e^{(\sum_{i=1}^{|f_{c,t}(X)|}f_{c,t}^i(X))}}$$

Methods	0.1	0.2	0.3	0.4	0.5
GMP	52.8	46.4	37.6	29.0	18.4
GAP	50.6	45.1	36.5	27.9	17.5
SAP	**60.1**	**54.6**	**45.1**	**34.3**	**22.4**

Reported CAS [28]	√	√	√
SSG (CAS w/ SSG)		√	√
ASSG (CAS w/ SSG w/ classifier)			√
Ave-mAP(%)	24.4	34.2	**43.5**

EXPERIMENT



Results on THUMOS'14 Dataset

Method	Label	0.1	0.2	0.3	0.4	0.5	0.6	0.7
Richard et al. [18]	strong	39.7	35.7	30.0	23.2	15.2	–	–
S-CNN [21]	strong	47.7	43.5	36.3	28.7	19.0	10.3	5.3
CDC [19]	strong	–	–	40.1	29.4	23.3	13.1	7.9
Gao et al. [6]	strong	54.0	50.9	44.1	34.9	25.6	19.1	9.9
Xu et al. [27]	strong	54.5	51.5	44.8	35.6	28.9	–	–
SSN [30]	strong	**66.0**	**59.4**	51.9	41.0	29.8	19.6	10.7
SSAD [13]	strong	50.1	47.8	43.0	35.0	24.6	–	–
TPC [29]	strong	–	–	44.1	37.1	28.2	20.6	12.7
TALNet [4]	strong	59.8	57.1	53.2	**48.5**	**42.8**	–	–
Alwasssel et al.[1]	strong	49.6	44.3	38.1	28.4	19.8	–	–
BSN [14]	strong	–	–	**53.5**	45.0	36.9	**33.8**	**20.8**
UntrimmedNet [24]	weak	44.4	37.7	28.2	21.1	13.7	–	–
Hide-and-Seek [23]	weak	36.4	27.8	19.5	12.7	6.8	–	–
Zhong et al. [31]	weak	45.8	39.0	31.1	22.5	15.9	–	–
AutoLoc [20]	weak	–	–	35.8	29.0	21.2	**13.4**	5.8
W-TALC [17]	weak	55.2	49.6	40.1	31.1	22.8	–	**7.6**
STPN [15]	weak	52.0	44.7	35.5	25.8	16.9	9.9	4.3
STAR [28]	weak	**68.8**	**60.0**	**48.7**	**34.7**	**23.0**	–	–
STPN-CAS w/ ASSG	weak	55.6	49.5	41.1	31.5	20.9	13.7	5.9
STAR-CAS w/ ASSG	weak	65.6	59.4	**50.4**	**38.7**	**25.4**	**15.0**	6.6