ISCAS 中国科学院软件研究所学术年会’2023
暨计算机科学国家重点实验室开放周

学术论文

基于原子对比的弱监督视频时刻定位

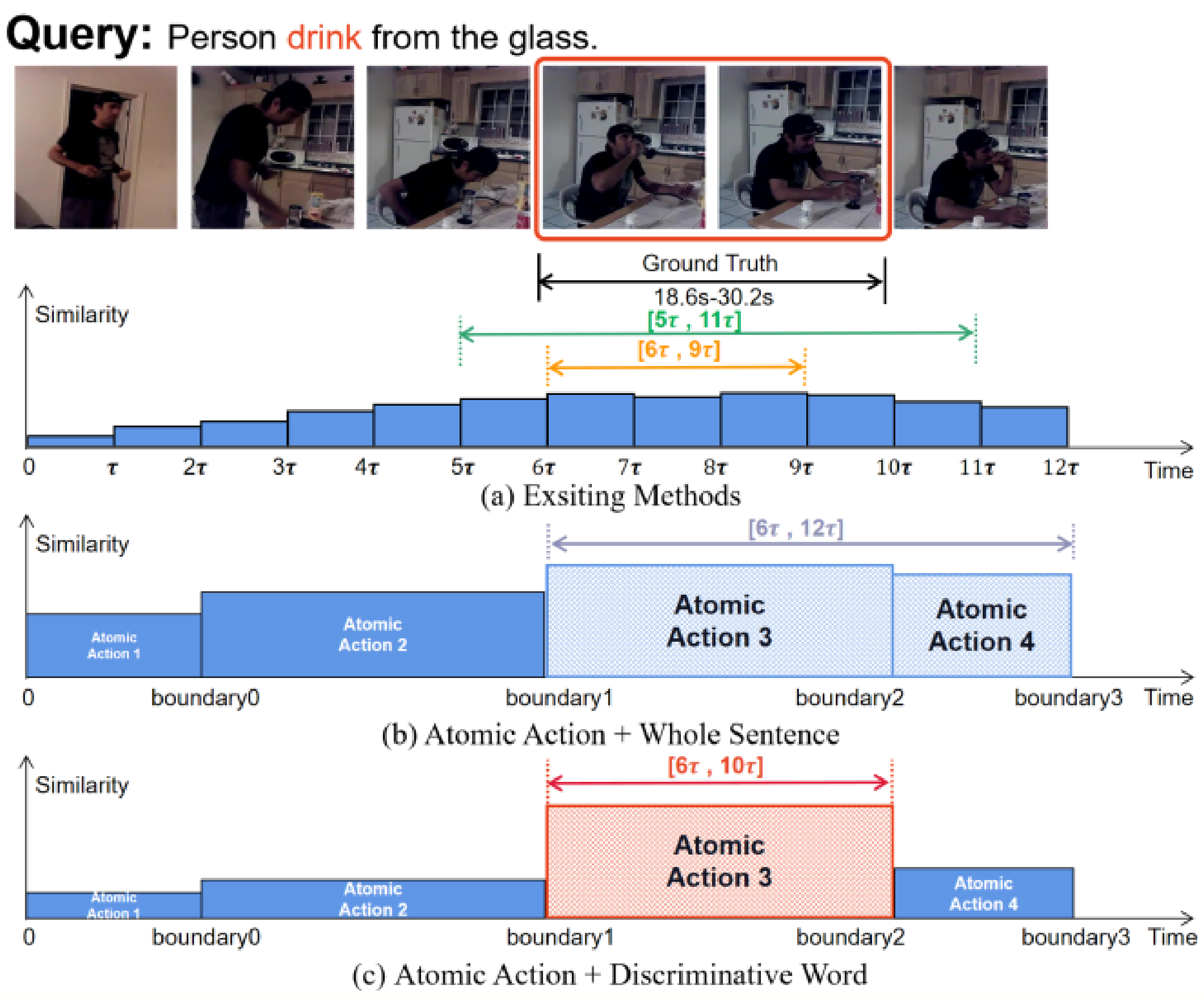
Atomic-action-based Contrastive Network for Weakly Supervised Temporal Language Grounding

闫一帆

电话：18511165056 邮箱：yifan2018@iscas.ac.cn

发表于ICME2023

任务描述：



(a) Exsiting Methods

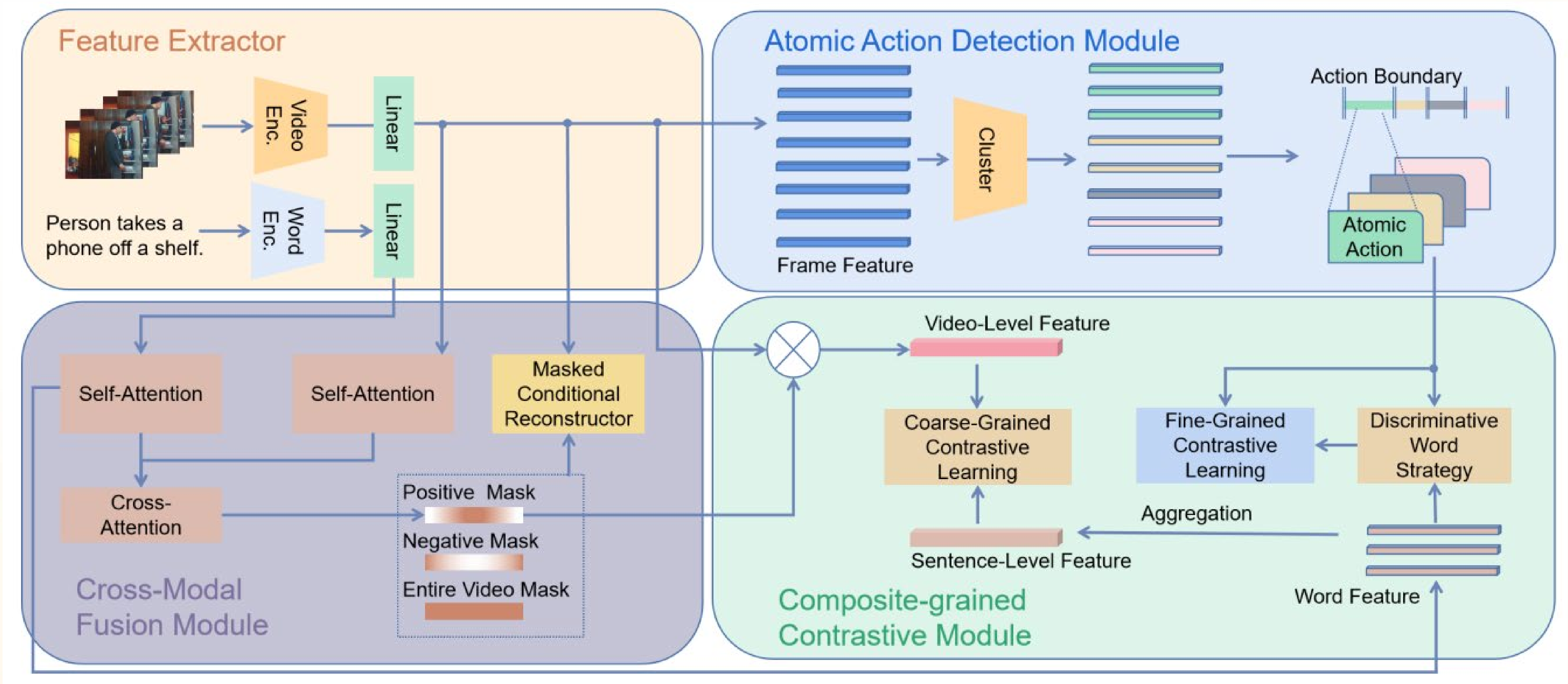
(b) Atomic Action + Whole Sentence

(c) Atomic Action + Discriminative Word

视频时刻定位任务，是通过给定一句自然语言描述的查询语句，在未剪裁的完整视频中确定该查询事件发生的始末帧，而弱监督的视频时刻定位任务则并未在训练数据中明确标注查询事件的始末帧，而仅提供查询-视频对。

本文方法：

一个视频中的事件可以在时序上确定性地分割为若干具有原子性的动作序列。一个动作的原子性情况一般被理解为一个动作在视觉上有准确的时刻边界。因此，本文将视频时刻定位任务中目标获取的事件始末帧转化为始末动作，以定位事件边界进行对比学习，从而获得更加稳定的定位边界。



性能指标：

ActivityNet数据集	IoU=0.1	IoU=0.3	IoU=0.5
DCCP	–	41.6%	23.2%
WS-DEC	62.71%	41.98%	23.34%
EC-SL	68.48%	44.29%	24.26%
MARN	–	47.01%	29.95%
SCN	71.48%	47.23%	29.22%
RTBPN	73.73%	49.77%	29.63%
WSLLN	75.4%	42.8%	22.7%
LCNet	78.58%	48.49%	26.33%
WSTAN	79.78%	52.45%	30.01%
CRM	**81.61%**	55.26%	32.19%
CNM	78.13%	55.68%	33.33%
Ours	78.78%	**57.66%**	**34.18%**

Charades-STA数据集	IoU=0.3	IoU=0.5	IoU=0.7
TGA	32.14%	19.94%	8.84%
SCN	42.96%	23.58%	9.97%
DCCP	–	29.8%	11.9%
WSTAN	43.39%	29.35%	12.28%
LoGAN	48.04%	31.74%	13.71%
MARN	48.55%	31.94%	14.81%
CRM	53.66%	34.76%	16.37%
LCNet	59.60%	**39.19%**	**18.87%**
RTBPN	60.04%	32.36%	13.24%
CNM	60.04%	35.15%	14.95%
Ours	**62.95%**	37.02%	15.26%

公开数据集上的实验证明，本文方法具有性能优越性。