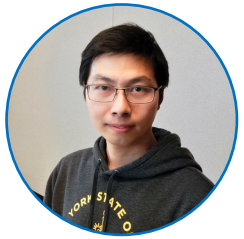
ACTION LAB

Machine Learning and Data Intensive Computing (Mining) LAB

CVPR JUNE 19-24 2022 NEW ORLEANS LOUISIANA

OpenTAL: Towards Open Set Temporal Action Localization

(CVPR 2022 Oral)



Wentao Bao

Qi Yu

Yu Kong

Golisano College of Computing and Information Sciences

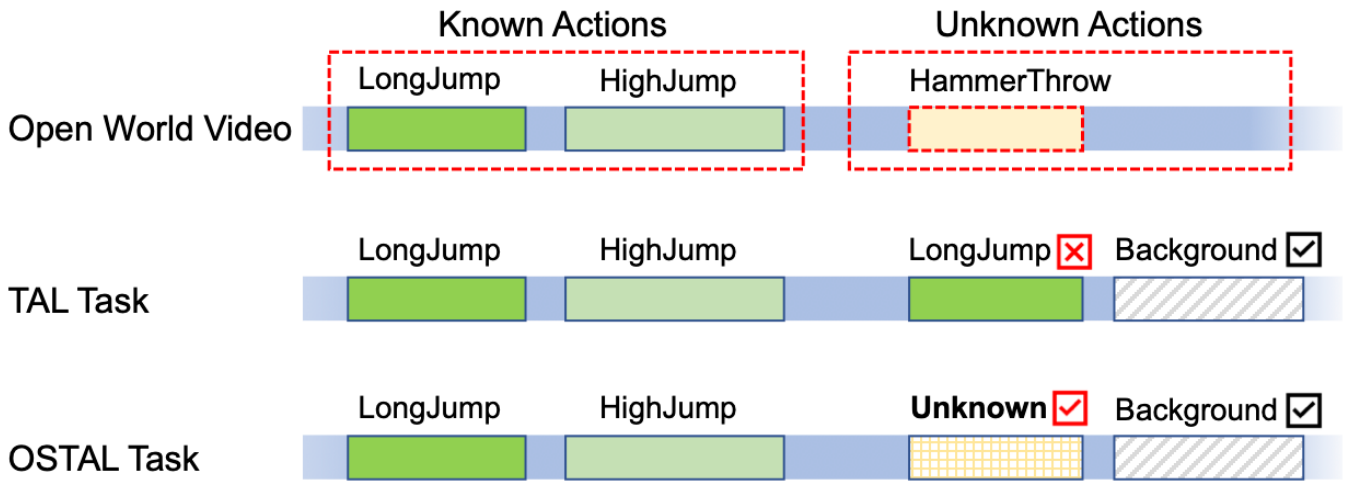
Rochester Institute of Technology (RIT)

Open Set Temporal Action Localization (OSTAL)

Problem Setup

Given an untrimmed video, a model needs to

- ❑ **Localize** and **recognize** actions/events from known classes and **reject** the unknown.
- ❑ Be trained with **known** classes (closed-set), tested with **known + unknown** classes (open-set).



Related Work

☐ Temporal Action Localization (TAL)

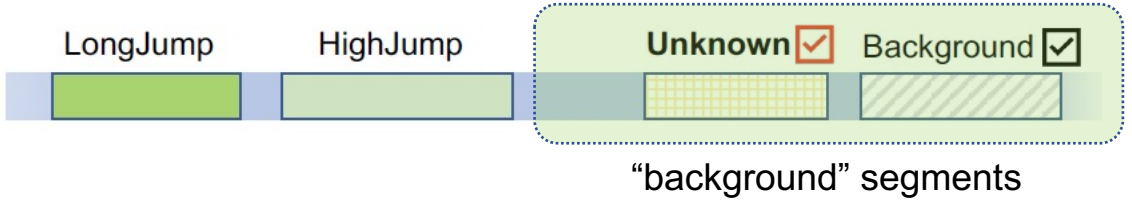
- **One-stage**: frame-level boundary and class prediction & global temporal reasoning.
  - FrameGlimpse (cvpr-16), SS-TAD (bmvc-17), GTAN (cvpr-19), **AFSD** (cvpr-21), …
- **Two-stage**: class-agnostic temporal proposals + classification & boundary refinement.
  - R-C3D (iccv-17), P-GCN (iccv-19), G-TAD (cvpr-20), CSA (iccv-21), …

☐ Open Set Recognition (OSR)

- **OSR for Image**: Discriminative / Generative / Prototypical / Uncertainty
  - W-SVM (tpami-14), OpenMax (cvpr16), GCPL (cvpr-18), C2AE (cvpr19), ARPL (tpami-21), OpenGAN (iccv-21),…
- **OSR for Video**: Temporal Reasoning
  - ODN (icme-18), BAR (neurips-18), **DEAR** (iccv-21)…
- **A Surge of OSR Applications:** Task-relevant Inducive Bias
  - 2D det (cvpr-21), 3D det (3dv-21), segmentation (iccv-21), audio-visual (iccv-19), few-shot (cvpr-20), adversarial defense (eccv-20), 6D pose (cvpr-22), anomaly (cvpr-22), tracking (cvpr-22)…

Our Motivations

Technical Challenges



- ❑ ***Background*** CANNOT be removed!
  - Provide context for temporal localization.
- ❑ ***Unknown*** actions CANNOT be removed!
  - Temporal annotations of unknown actions are NOT available.
- ❑ ***Semi-supervision***: unknown actions are MIXED with pure background frames

Naïve Solutions

- ❑ Regard the ***Unknown*** as the (K+1)-th class?
  - Nontrivial to generate training data without temporal annotations.
- ❑ **Manually label** the temporal boundary of the unknown?
  - Meaningless setting, vague semantics of the unknown.

Proposed OpenTAL Model

	Known Action	Unknown Action	Background
uncertainty (u)	↓	↑	↑
actionness (a)	↑	↑	↓

Framework

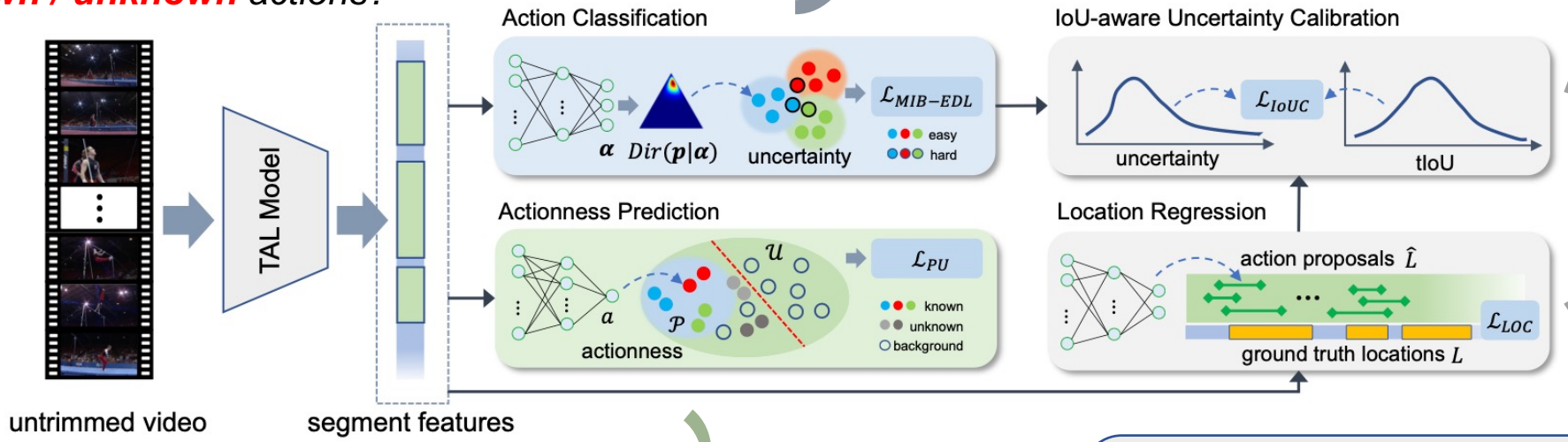
$$u_i = K/S_i$$

$$w_{\hat{l}_i,l_i} = \max\left(\gamma, \mathrm{IoU}(\hat{l}_i, l_i)\right)$$

$$\mathcal{L}_{\text{MIB-EDL}} = \frac{1}{N}\sum_{i=1}^{N} \tilde{\omega}_i^{(t)} \mathcal{L}_{\text{EDL}}^{(i)}(\boldsymbol{\alpha}_i). \qquad \mathcal{L}_{\text{EDL}}^{(i)}(\boldsymbol{\alpha}_i) = \sum_{j=1}^{K} t_{ij}(\log(S_i) - \log(\alpha_{ij})),$$

$$\mathcal{L}_{\text{IoUC}}^{(i)}(\hat{l}_i, u_i) = -w_{\hat{l}_i,l_i}\log(1-u_i) - (1-w_{\hat{l}_i,l_i})\log(u_i)$$

*"Are they **known / unknown** actions?"*



$$\mathcal{L}_{\text{ACT}}(\hat{\mathcal{P}}, \hat{\mathcal{N}}) = -\frac{1}{|\hat{\mathcal{P}}|}\sum_{\hat{a}_i \in \hat{\mathcal{P}}} \log \hat{a}_i - \frac{1}{|\hat{\mathcal{N}}|}\sum_{\hat{a}_i \in \hat{\mathcal{N}}} \log(1-\hat{a}_i).$$

$$\hat{\mathcal{N}} = \{\hat{a}_i | \hat{a}_i \in sort(\hat{\mathcal{U}})_{1,\ldots,M}\}$$ *"Foreground / **background**?"*

$$\begin{cases} \mathcal{L}_{\text{LOC}}(\{\hat{l}_i\}) = \dfrac{1}{N_C}\displaystyle\sum_i \mathbb{I}[y_i \geq 1]\left(1 - \frac{|\hat{l}_i \cap l_i|}{|\hat{l}_i \cup l_i|}\right) \\ \mathcal{L}_{\text{LOC}}(\{\hat{\delta}_i\}) = \dfrac{1}{N_R}\displaystyle\sum_i \mathbb{I}[y_i \geq 1](|\hat{\delta}_i - \delta_i|), \end{cases}$$

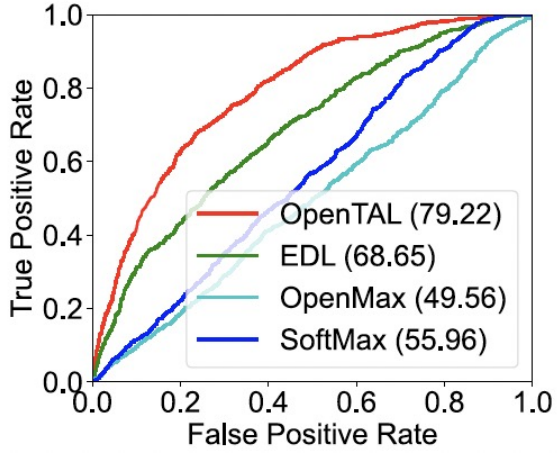
*"**Where** are these actions?"*

Experimental Results

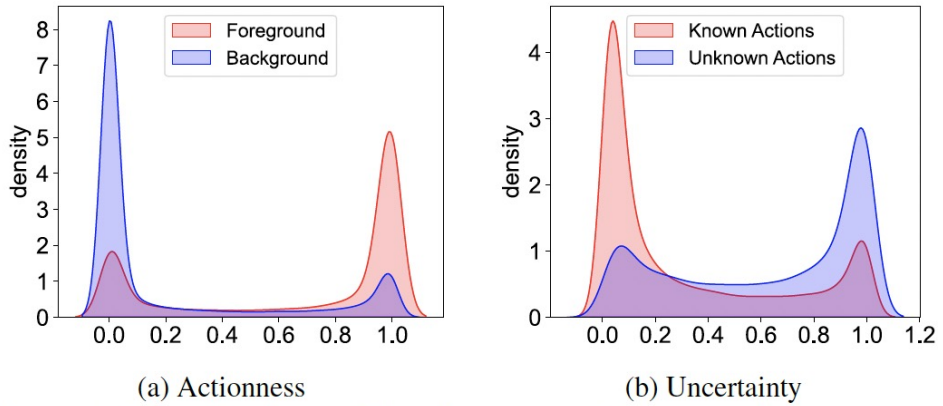
Compare with SOTA

Table 2. **OSTAL Results (%).** Models trained on the THUMOS14 closed set are tested on the open sets by including the unknown classes from THUMOS14 and ActivityNet1.3, respectively. The mAP is provided as the reference of the TAL results on THUMOS14 closed set.

Methods	THUMOS14 as the Unknown				ActivityNet1.3 as the Unknown				mAP
	FAR@95 (↓)	AUROC	AUPR	OSDR	FAR@95 (↓)	AUROC	AUPR	OSDR	
SoftMax	85.58	54.70	31.85	23.40	85.05	56.97	53.54	27.63	55.81
OpenMax [6]	90.34	53.26	33.17	13.66	91.36	51.24	54.88	15.73	36.36
EDL [4]	81.42	64.05	40.05	36.26	84.01	62.82	53.97	38.56	52.24
OpenTAL	**70.96**	**78.33**	**58.62**	**42.91**	**63.11**	**82.97**	**80.41**	**50.49**	55.02



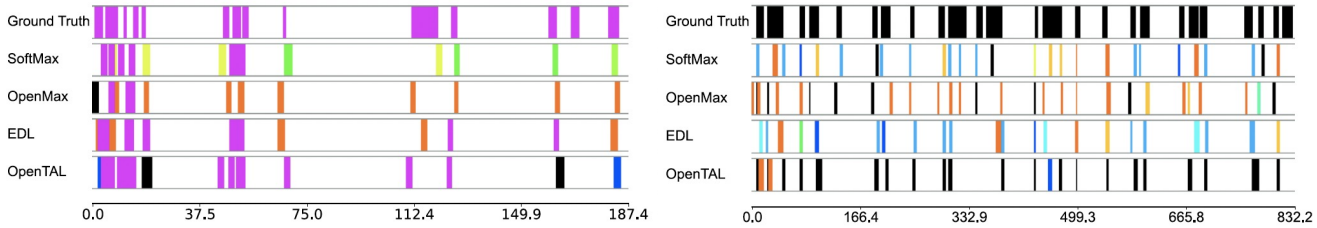
Model Analysis



(a) Actionness (b) Uncertainty

Figure 4. **Distributions of Actionness and Uncertainty.** The two

Visualization



Accurate Closed-set TAL Accurate Open-set TAL

Welcome to our poster session for more discussion!

THANKS

Paper: https://arxiv.org/pdf/2203.05114.pdf

Code: https://github.com/Cogito2012/OpenTAL

Paper

Code

Feel free to contact me via wtbao2018@gmail.com.