TSTM: Temporal Segmentation for Task-relevant Mask in Visual Reinforcement Learning Generalization

Weicheng Du[1], Wenjia Meng[1*], Zhengzhe Zhang[1], Yilong Yin[1], Xiankai Lu[1,2]
[1]School of Software, Shandong University [2]School of Mathematics and Computer Science, Quanzhou Normal University

Time: Jun 6th Afternoon Poster Session: 4
Session Order: 613 Poster #30654

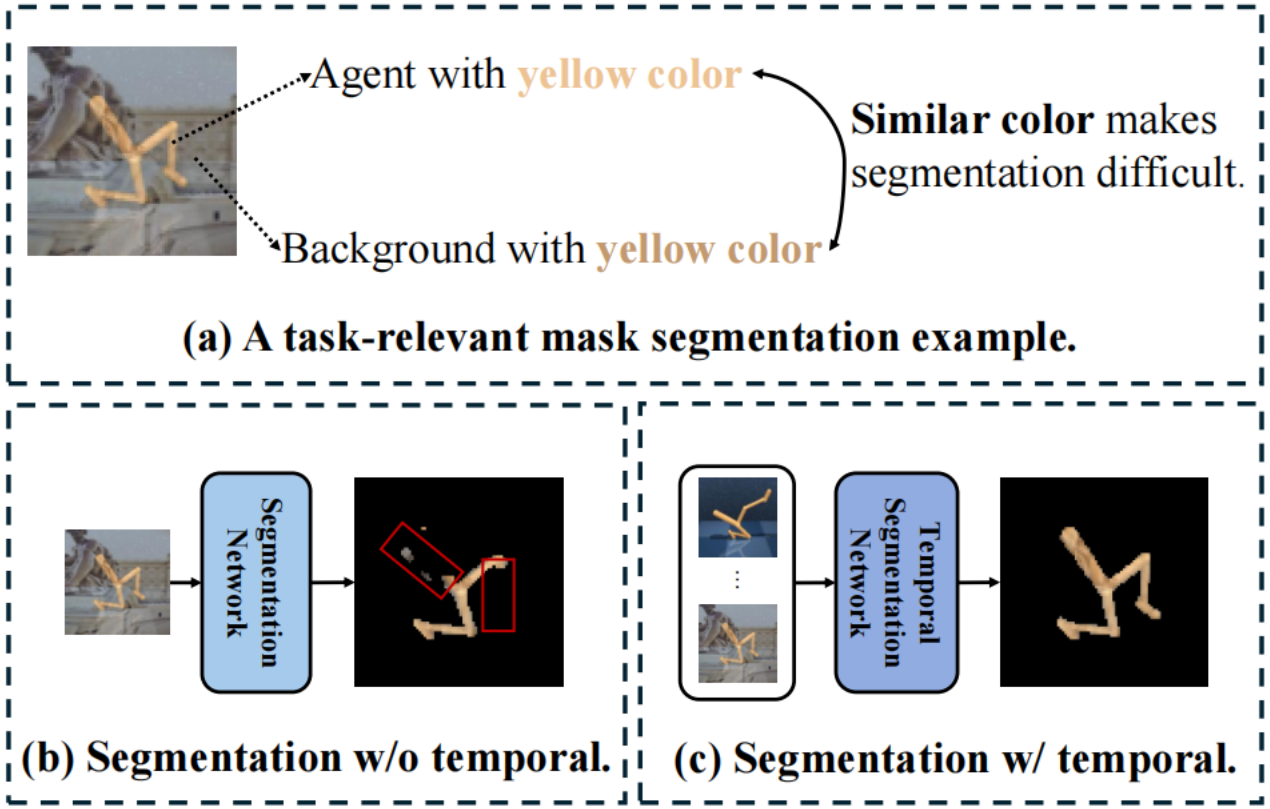
By: Weicheng Du

CVPR JUNE 3-7, 2026 DENVER COLORADO

Outline

- Motivation
- TSTM Framework
- Temporal Segmentation Network
- Optimization Objectives
- Experiments

Motivation



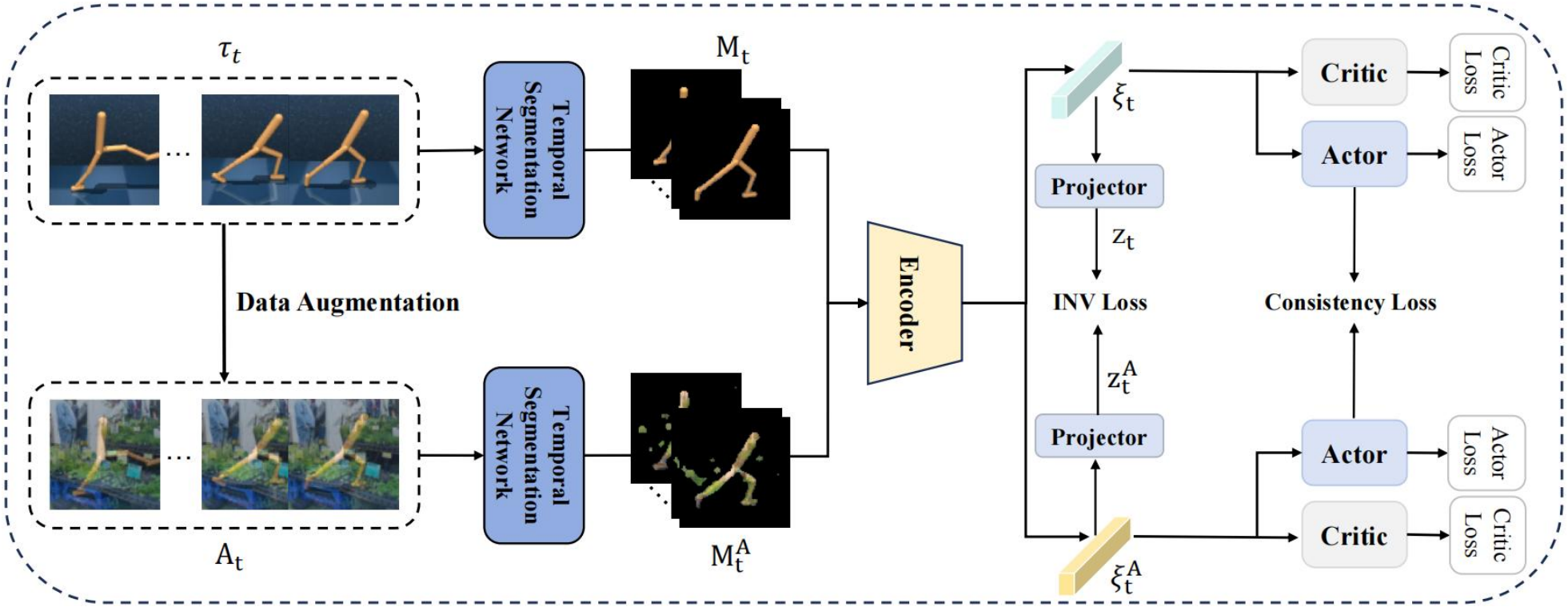
1. **Single-frame masking is unreliable in visual RL.**
2. **Ambiguous backgrounds can confuse task-relevant segmentation.**
3. **Temporal cues help distinguish agents from distracting visual content.**

Goal: learn more reliable task-relevant masks from observation sequences.

[1]Du, Weicheng, et al. "TSTM: Temporal Segmentation for Task-relevant Mask in Visual Reinforcement Learning Generalization." CVPR 2026.

https://github.com/sduwcdd/tstm

TSTM Framework



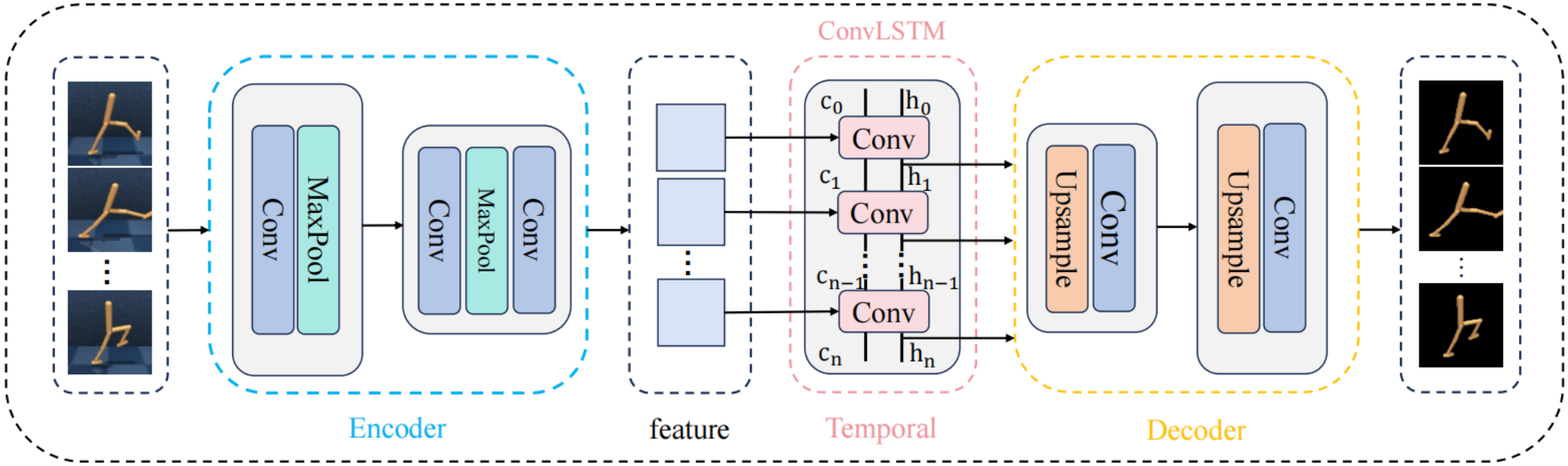
key components:

- **Temporal segmentation**
- **Invariant learning**
- **Policy consistency**

[1]Du, Weicheng, et al. "TSTM: Temporal Segmentation for Task-relevant Mask in Visual Reinforcement Learning Generalization." CVPR 2026.

https://github.com/sduwcdd/tstm

Temporal Segmentation Network



key components:

- **Encoder:** extracts spatial features from sequential observations
- **ConvLSTM:** captures temporal dependencies across frames
- **Decoder:** predicts more reliable task-relevant masks

Temporal modeling improves segmentation reliability under visual ambiguity.

[1]Du, Weicheng, et al. "TSTM: Temporal Segmentation for Task-relevant Mask in Visual Reinforcement Learning Generalization." CVPR 2026.

https://github.com/sduwcdd/tstm

Optimization Objectives

Teacher / Student Segmentation

- We first train a teacher segmentation network and distill it into a lightweight student for efficient inference.

$$\mathcal{L}_{\text{BCE}} = -\frac{1}{H \times W}\left(y_i \odot \log u_i + (1-y_i) \odot \log(1-u_i)\right)$$

$$\mathcal{L}_{\text{Dice}} = 1 - \frac{2\langle y_i, u_i \rangle + \epsilon}{\|y_i\|_1 + \|u_i\|_1 + \epsilon}$$

Invariant Representation Learning

- Align masked representations of original and augmented observation sequences.

$$\mathcal{L}_{\text{mse}} = \|z_t - z_t^A\|_2^2$$

$$\mathcal{L}_{\text{var}} = \frac{1}{d}\sum_{i=1}^{a}\Big[\max(0, \Gamma - \sqrt{\text{Var}(z_{t,i}) + \eta}) + \max(0, \Gamma - \sqrt{\text{Var}(z_{t,i}^A) + \eta})\Big]$$

$$\mathcal{L}_{\text{cov}} = \frac{1}{d}\sum_{i \neq j}\left[C(z_t)_{ij}^2 + C(z_t^A)_{ij}^2\right]$$

$$\mathcal{L}_{\text{INV}} = \lambda\mathcal{L}_{\text{mse}} + \mu\mathcal{L}_{\text{var}} + \rho\mathcal{L}_{\text{cov}}$$

Actor Update with Policy Consistency

- The actor is optimized with standard SAC objective plus a policy consistency term between original and augmented masked observations.

$$\mathcal{L}_\pi = \mathbb{E}_{s_t}\left[\alpha \log \pi_\psi(a_t \mid s_t) - Q_\theta(s_t, a_t)\right]$$

$$\mathcal{L}_{\text{PC}} = \mathbb{E}_{\xi_t \sim \mathcal{B}}\left[D_{\text{KL}}\left(\text{sg}(\pi_\psi(\cdot \mid \xi_t)) \,\middle\|\, \pi_\psi(\cdot \mid \xi_t^A)\right)\right]$$

$$\mathcal{L}_A = \mathcal{L}_\pi + \varsigma\,\mathcal{L}_{\text{PC}}$$

INV loss acts in representation space, while policy consistency is added to the actor objective.

[1]Du, Weicheng, et al. "TSTM: Temporal Segmentation for Task-relevant Mask in Visual Reinforcement Learning Generalization." CVPR 2026.

https://github.com/sduwcdd/tstm

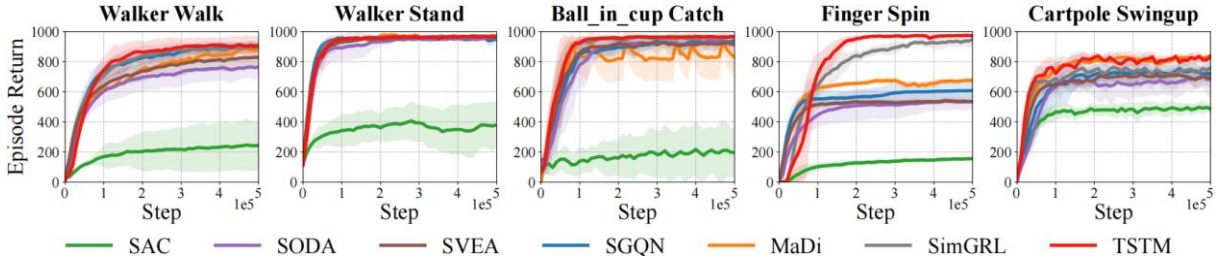
Experiments

TSTM achieves the best average return on DMC-GB benchmark.

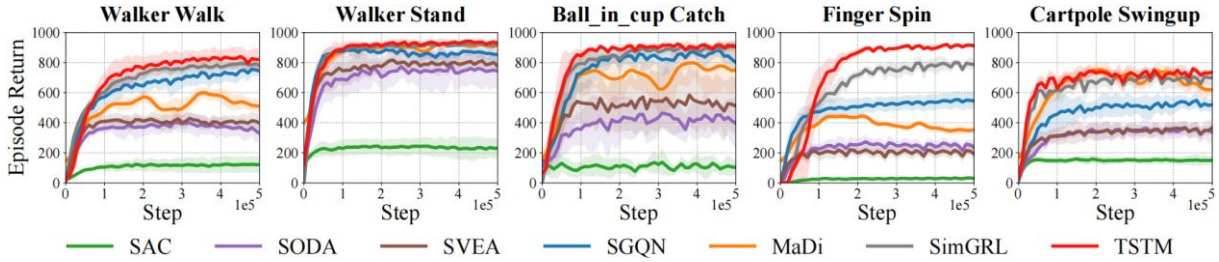
	DMC-GB [21]	SAC [74]	SODA [70]	SVEA [21]	SGQN [25]	MaDi [26]	SimGRL [75]	TSTM (Ours)
Video easy	Walker Walk	245±165	771±66	828±66	910±24	895±24	910±21	**912±42**
	Walker Stand	389±131	965±7	966±5	955±12	967±3	**973±4**	968±3
	Ball_in_cup Catch	192±157	939±10	908±55	950±24	807±144	964±7	**969±2**
	Finger Spin	157±8	535±53	537±11	610±61	679±17	957±16	**971±12**
	Cartpole Swingup	474±26	678±120	684±74	717±35	848±6	775±60	**851±19**
	average	291	778	785	828	839	916	**934**
Video hard	Walker Walk	122±47	312±32	385±63	739±21	504±33	773±31	**821±36**
	Walker Stand	231±57	736±132	747±43	851±24	920±14	**932±17**	923±25
	Ball_in_cup Catch	101±37	381±163	498±174	782±57	758±135	902±19	**903±27**
	Finger Spin	26±6	222±48	174±39	541±53	358±25	784±10	**906±10**
	Cartpole Swingup	153±22	328±79	381±28	526±42	619±24	696±39	**741±22**
	average	127	396	437	688	632	817	**859**

TSTM consistently outperforms prior visual RL baselines on most tasks.

Learning curves on DMC-GB under the video easy setting



Learning curves on DMC-GB under the video hard setting



[1]Du, Weicheng, et al. "TSTM: Temporal Segmentation for Task-relevant Mask in Visual Reinforcement Learning Generalization." CVPR 2026.

https://github.com/sduwcdd/tstm

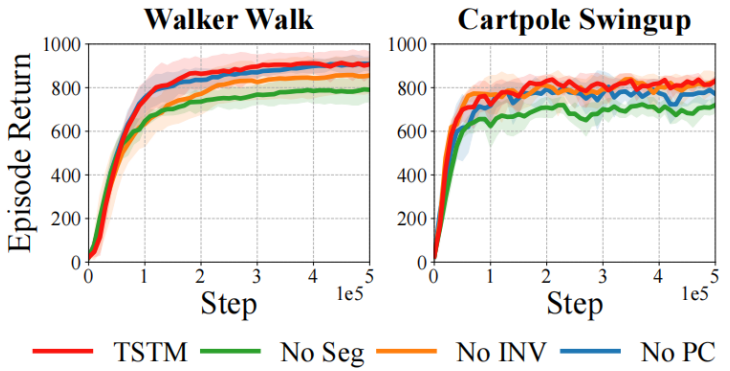
Experiments

Ablation Study

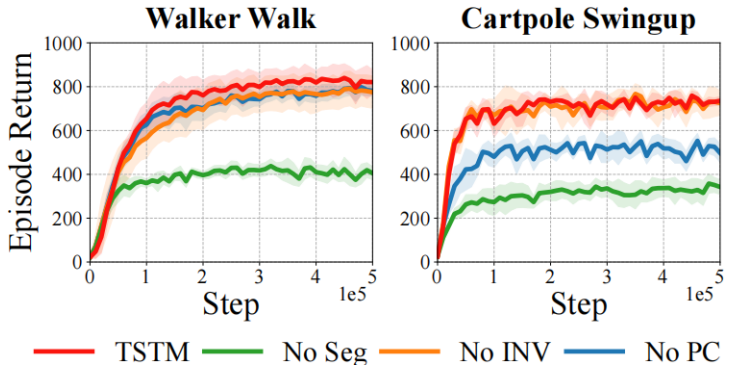
- **Ablation results on DMC-GB.**

	Method	Walker Walk	Cartpole Swingup
Video Easy	TSTM	**912±42**	**851±19**
	No Seg	788±47	733±42
	No INV	859±78	834±29
	No PC	908±34	752±120
Video Hard	TSTM	**821±36**	**741±22**
	No Seg	391±25	333±33
	No INV	770±73	720±44
	No PC	777±22	469±44

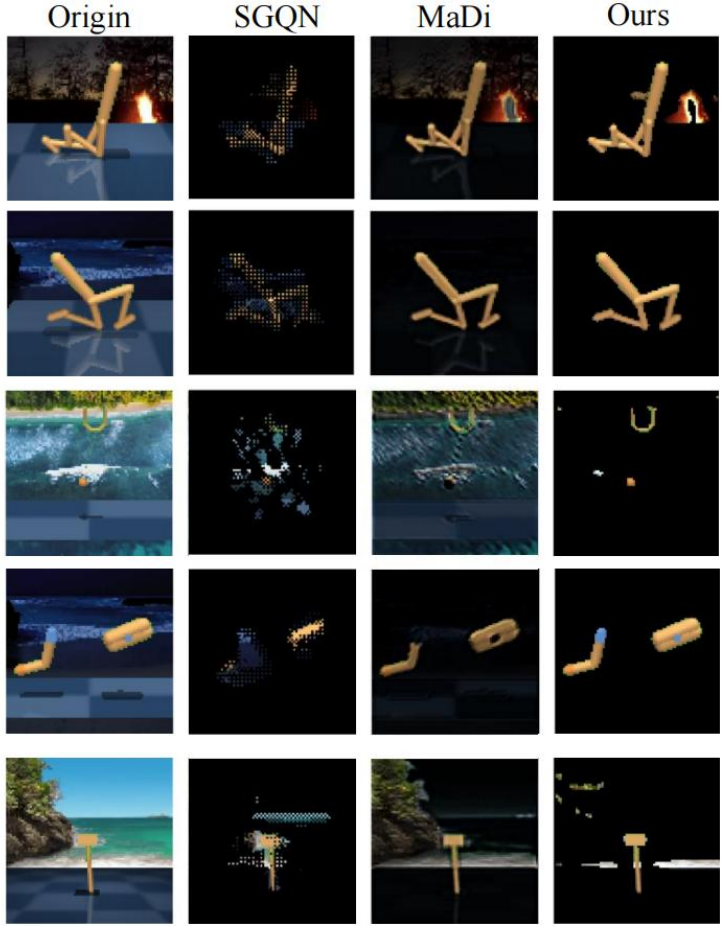
- **Ablation curves on DMC-GB (*video easy*).**



- **Ablation curves on DMC-GB (*video hard*).**



Visualization



[1]Du, Weicheng, et al. "TSTM: Temporal Segmentation for Task-relevant Mask in Visual Reinforcement Learning Generalization." CVPR 2026.

https://github.com/sduwcdd/tstm

Thanks !

Time: Jun 6th Afternoon Poster Session: 4

Session Order: 613 Poster #30654