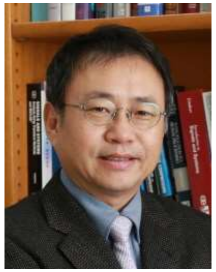
A Simple Framework for Generalization in Visual RL under Dynamic Scene Perturbations






Wonil Song
Yonsei University
Seoul, South Korea

Hyesong Choi
Ewha Womans University
Seoul, South Korea

Kwanghoon Sohn
Yonsei University
Seoul, South Korea

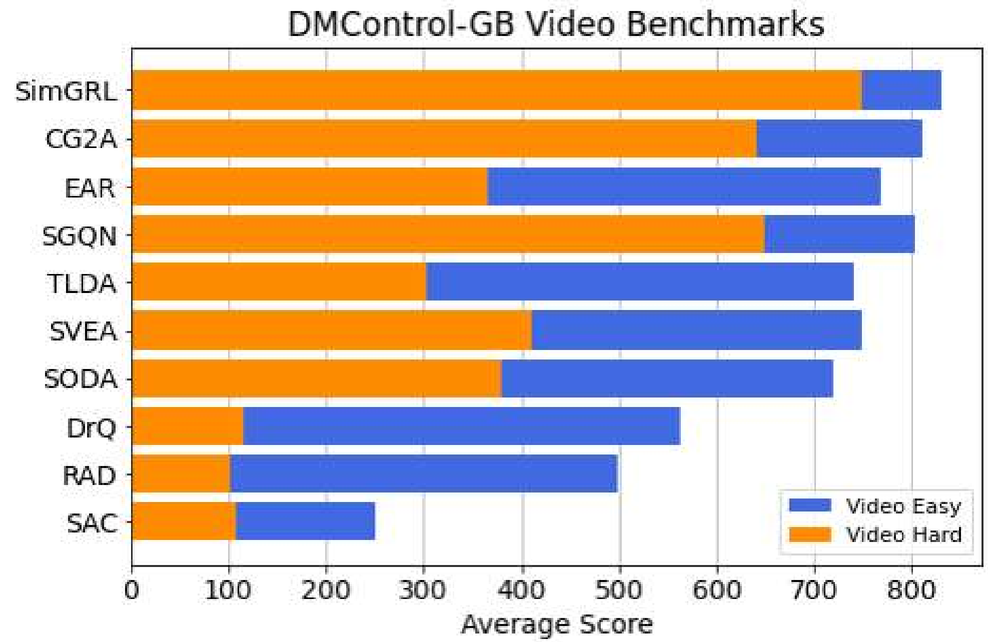
Dongbo Min
Ewha Womans University
Seoul, South Korea

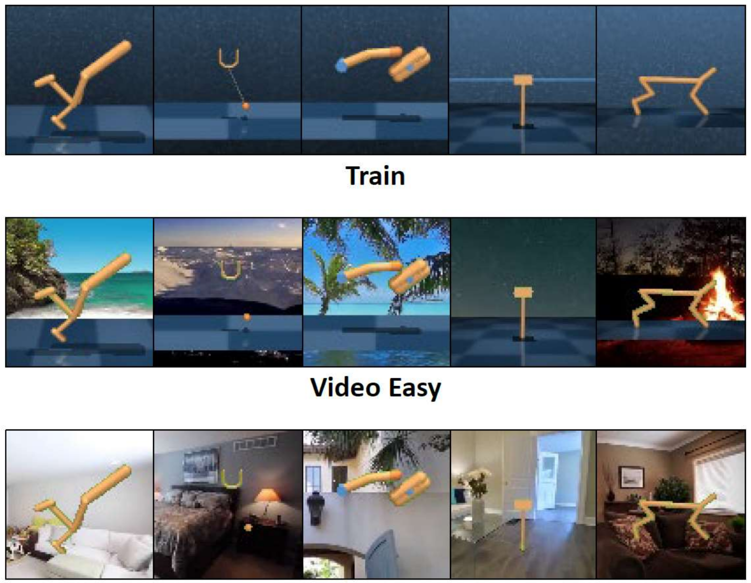
NeurIPS 2024

Motivation

Performance Degradation in Challenging Environments

- Existing algorithms for the generalization of *vision-based reinforcement learning (RL)* exhibit significant performance degradation in challenging environments like Video Hard in the DMControl-GB[1,2].
- Our proposed SimGRL demonstrates robust performance across all benchmarks.





[1] "Deepmind control suite." *arXiv* (2018).
[2] "Generalization in reinforcement learning by soft data augmentation." *ICRA* (2021).

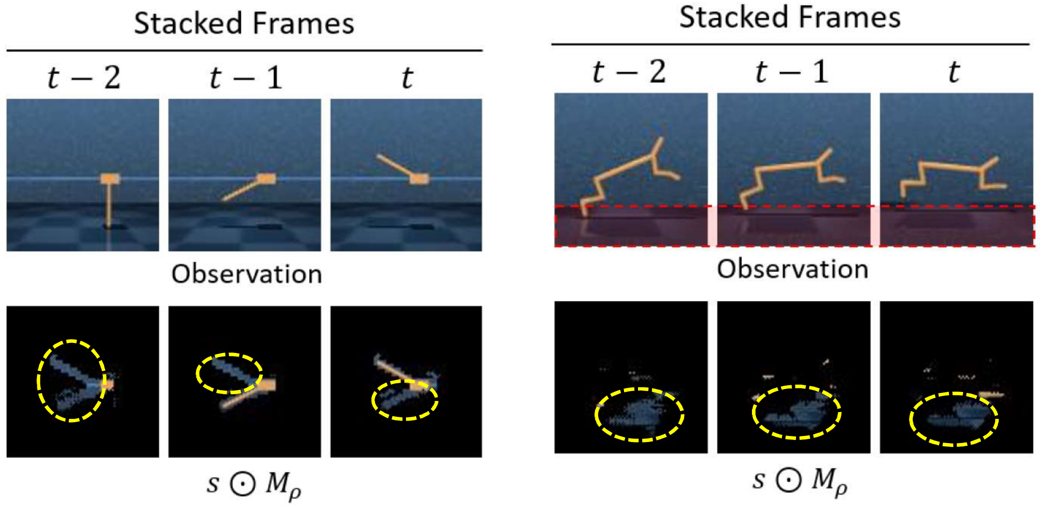
Problem Statement

Core Problems Causing Overfitting

Existing methods were vulnerable to the following issues :

- *Imbalanced saliency.*
- *Observational overfitting*[1].

M_ρ: Attribution mask obtained from the binarization by ρ-quantile of a gradient-based attribution map.



Imbalanced Saliency **Observational Overfitting**

Causing overfitting to training environments

[1] "Observational overfitting in reinforcement learning." *ICLR* (2020).

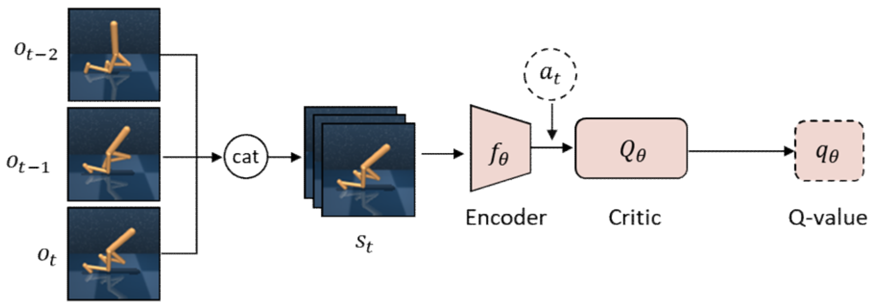
Problem Statement

Conventional Practices in Visual RL for Generalization

- *Image-level frame stack* used to encode temporal information.
- *Data augmentation uniformly applied across consecutive frames* used to learn robust representations (+ optionally representation learning).

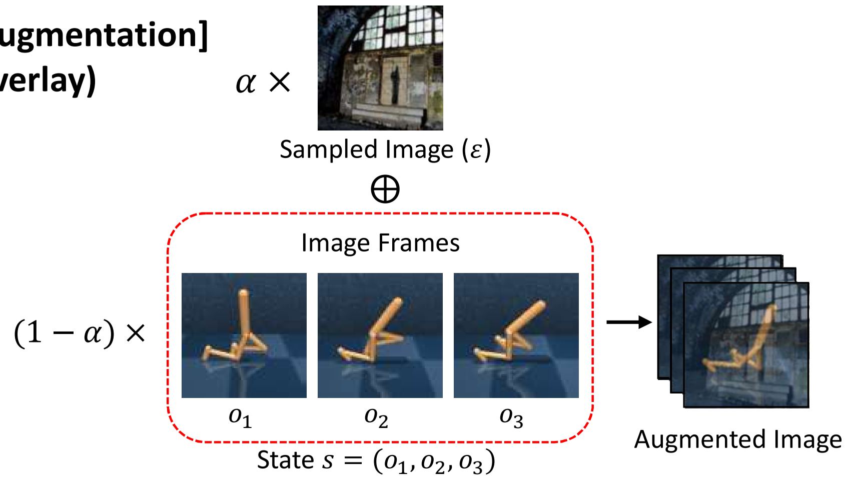
Insufficient!

[Image-level Frame Stack]



$$q_\theta(s_t, a_t) = Q_\theta(f_\theta([o_{t-2}, o_{t-1}, o_t]), a_t), \quad s_t = (o_{t-2}, o_{t-1}, o_t)$$

[Example of data augmentation]
(Random overlay)



Summary

We propose two simple regularization strategies to mitigate the problems.

1. **Architectural modification**
   - We propose to modify the structure of the image encoder.

2. **Data augmentation**
   - We propose a more proper data augmentation.

Proposed Method

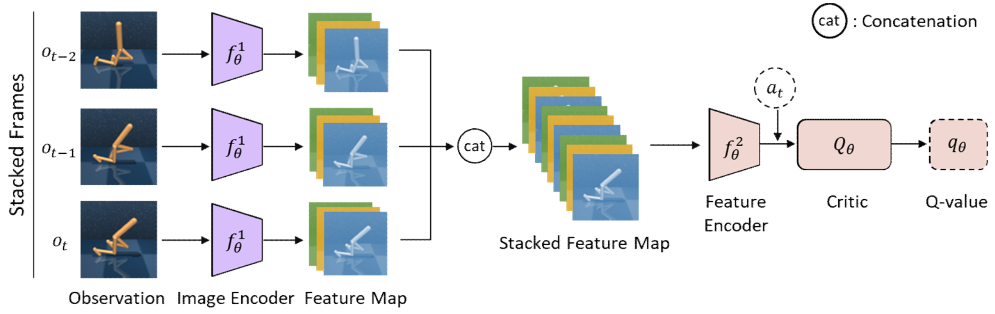
1. Feature-Level Frame Stack

- Modifying the structure of the encoder to enable an RL agent to separately focus on important features for each frame.

2. Shifted Random Overlay Augmentation

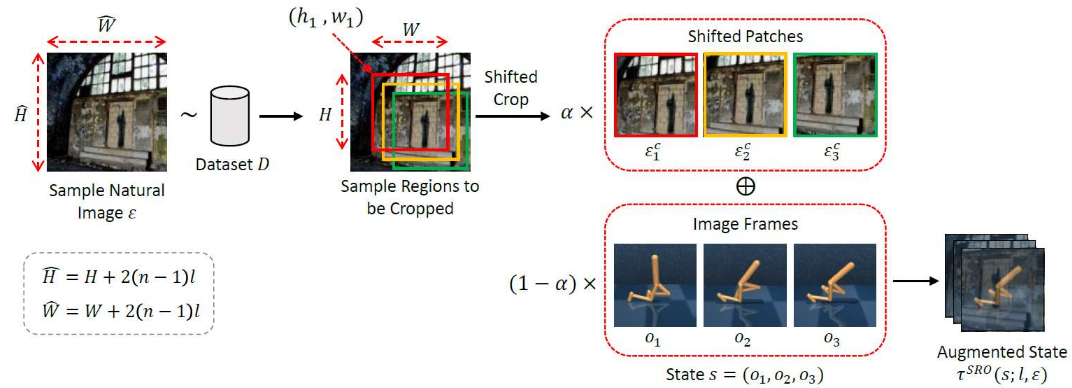
- Modifying random overlay[1] augmentation to enable an RL agent to be robust to dynamic background distractions and observational overfitting.

[Feature-Level Frame Stack]



$$q_\theta(s_t, a_t) = Q_\theta(f_\theta^2([f_\theta^1(o_{t-2}), f_\theta^1(o_{t-1}), f_\theta^1(o_t)]), a_t), \quad s_t = (o_{t-2}, o_{t-1}, o_t)$$

[Shifted Random Overlay Augmentation]

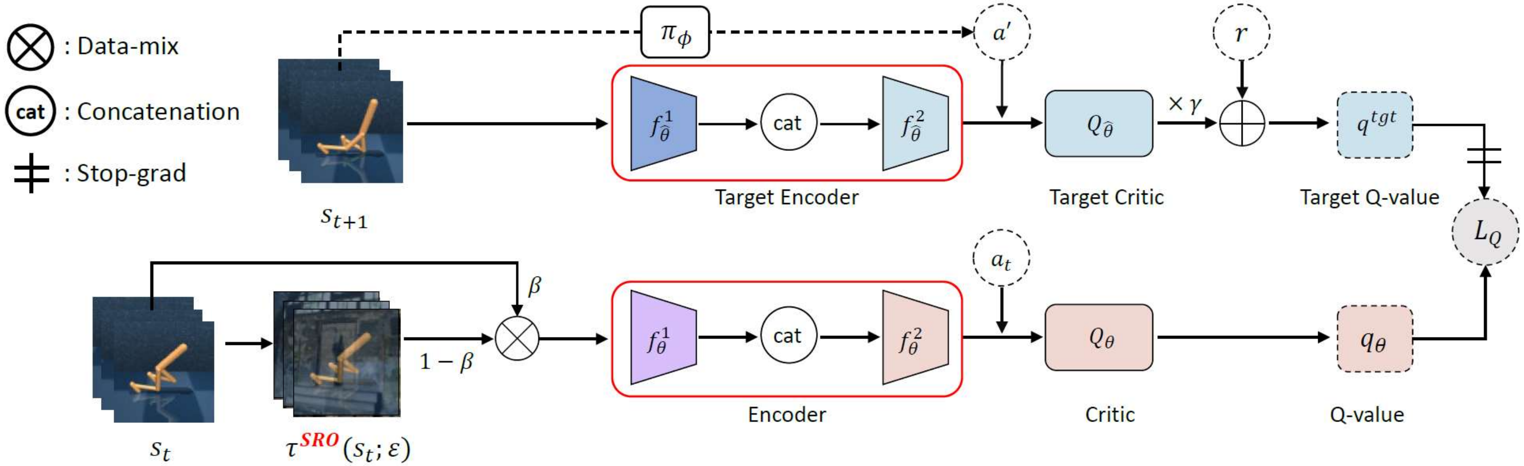


[1] "Generalization in reinforcement learning by soft data augmentation." *ICRA* (2021).

Proposed Method

SimGRL – A Simple Framework for Generalization in Visual RL under Dynamic Scene Perturbations

- Integrating the two proposed regularizations.
- Adopting the SVEA[1] algorithm for our baseline.



$$\mathcal{L}_Q(\theta) = \mathbb{E}_{(s_t,a_t,r_t,s_{t+1})\sim\mathcal{B}}[\beta(q_\theta(s_t,a_t) - q^{tgt})^2 + (1-\beta)(q_\theta(\tau^{SRO}(s_t;l,\varepsilon),a_t) - q^{tgt})^2]$$

where $q_\theta(s_t,a_t) = Q_\theta(f_\theta^2([f_\theta^1(o_{t-2}), f_\theta^1(o_{t-1}), f_\theta^1(o_t)]),a_t), \quad s_t = (o_{t-2}, o_{t-1}, o_t)$

[1] "Stabilizing deep q-learning with convnets and vision transformers under data augmentation." *NeurIPS* (2021).

Proposed Method

NEURAL INFORMATION PROCESSING SYSTEMS

Task-IDentification (TID) Metrics

- Measuring ***quantitatively*** the ability for the model to identify task-relevant objects.
- Providing a useful tool to analyze the problems.

TID Score

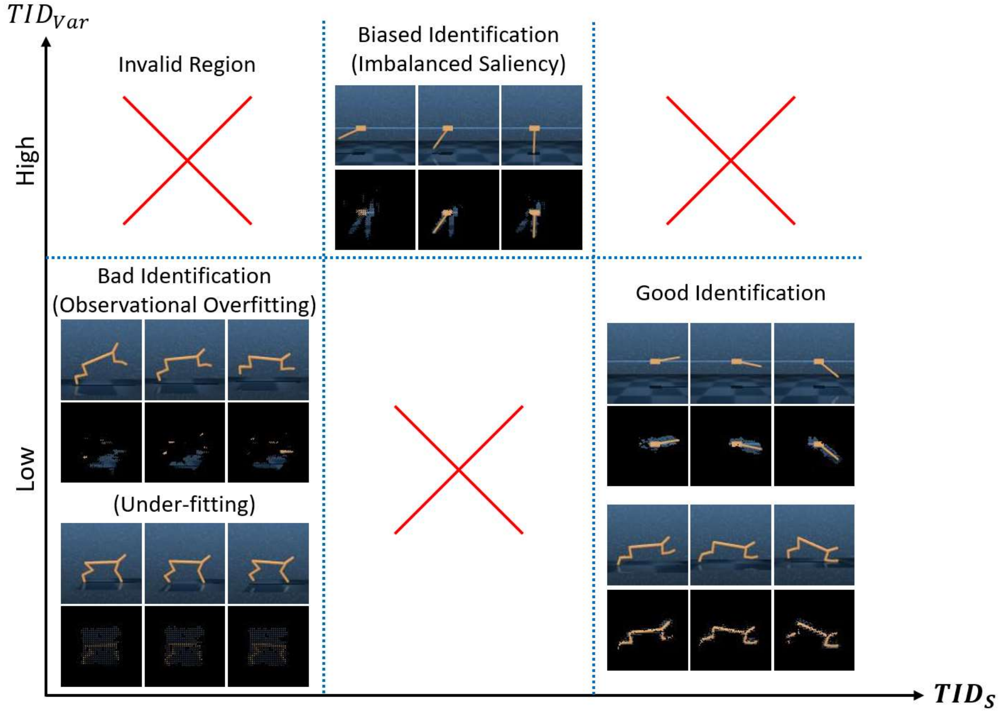
$$TID_S = \sqrt{\frac{N_{obj_M}}{N_{obj}} \times \frac{N_{obj_M}}{N_M}} = \sqrt{\frac{(N_{obj_M})^2}{N_{obj} \times N_M}}$$

Where,
N_{obj} : Number of task object's pixels in input images.
N_M : Number of pixels in attribution masks M_ρ.
$N_{ob\ M}$: Number of task object's pixels included in M_ρ.

TID Variance

$$TID_{Var} = Var[100 \times (TID_S^1, TID_S^2, ..., TID_S^n)]$$

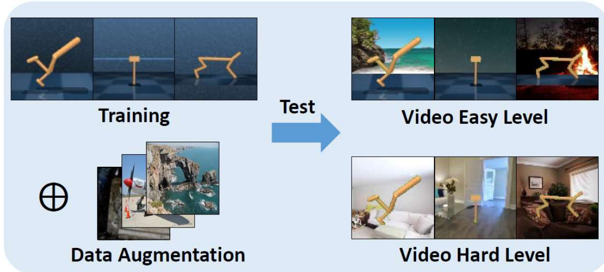
Where,
TID_S^i : Individually computed TID scores at each frame.



Experiments

Experiment Results on DMControl-GB Benchmarks

- We evaluated zero-shot test performances for video environments of DMControl-GB.
- Superior performance in Video Easy.
  (Reaching saturated performance by existing methods)
- *Significant performance improvement in Video Hard.*
  (Not yet reaching saturated performance by existing methods)

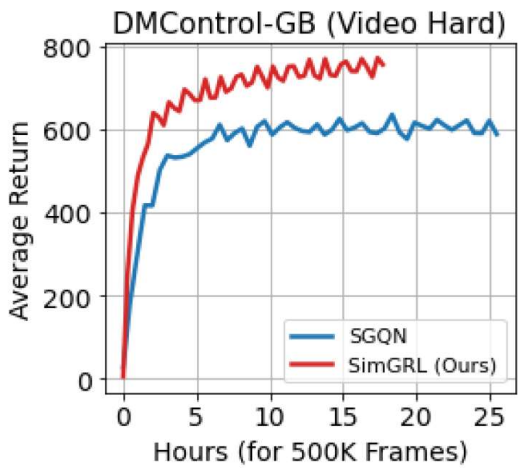


	DMControl-GB	SAC	RAD	DrQ	SODA	SVEA	TLDA	SGQN	EAR	CG2A	SimGRL	Δ
Video Easy	Walker, Walk	245±165	608±92	747±21	768±38	819±71	868±63	910±24	913±38	**918±20**	910±21	-8 (0.8%)
	Walker, Stand	389±131	879±64	926±30	955±13	961±8	973±6	955±9	970±23	968±6	**973±4**	0
	Ball In Cup, Catch	192±157	363±158	380±188	875±56	871±106	855±56	950±24	911±40	963±28	**964±7**	+1 (0.1%)
	Cartpole, Swingup	398±60	473±54	459±81	758±62	782±27	671±57	761±28	762±88	788±24	**838±35**	+50 (6%)
	Finger, Spin	206±169	516±113	599±62	695±97	808±33	744±18	956±28	717±51	912±69	**983±2**	+27 (3%)
	Cheetah, Run	73±18	153±7	270±16	268±10	251±17	**336±57**	289±35	334±56	314±49	317±16	-19 (6%)
Video Hard	Walker, Walk	122±47	80±10	121±52	312±32	385±63	292±133	739±21	383±59	687±18	**773±31**	+34 (5%)
	Walker, Stand	231±57	229±45	252±57	771±83	834±46	595±56	851±24	744±62	895±35	**932±17**	+37 (5%)
	Ball In Cup, Catch	101±37	98±40	100±40	327±100	403±174	304±58	782±57	320±48	806±44	**902±19**	+96 (12%)
	Cartpole, Swingup	158±17	152±29	136±29	429±64	393±45	308±44	544±43	375±37	472±24	**727±23**	+183 (34%)
	Finger, Spin	13±10	39±20	38±13	302±41	335±58	256±25	822±24	277±62	819±38	**864±12**	+42 (5%)
	Cheetah, Run	75±14	21±9	49±13	130±24	112±12	67±23	157±69	91±46	168±16	**301±7**	+133 (79%)

Experiments

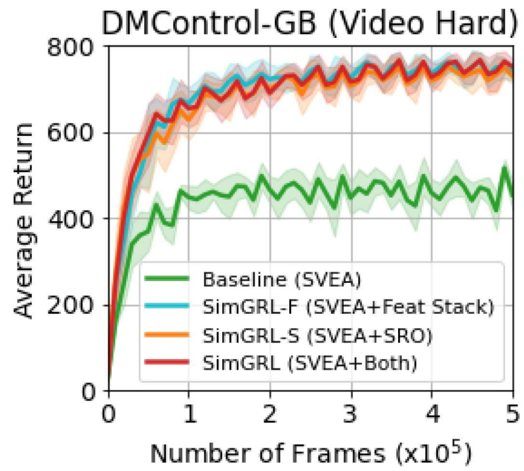
Computational Efficiency

- Thanks to the lack of any additional losses or networks, SimGRL is much more efficient than the previous SOTA SGQN[1].



Ablation Study

- Each regularization leads to remarkable performance improvements over the baseline SVEA[2].

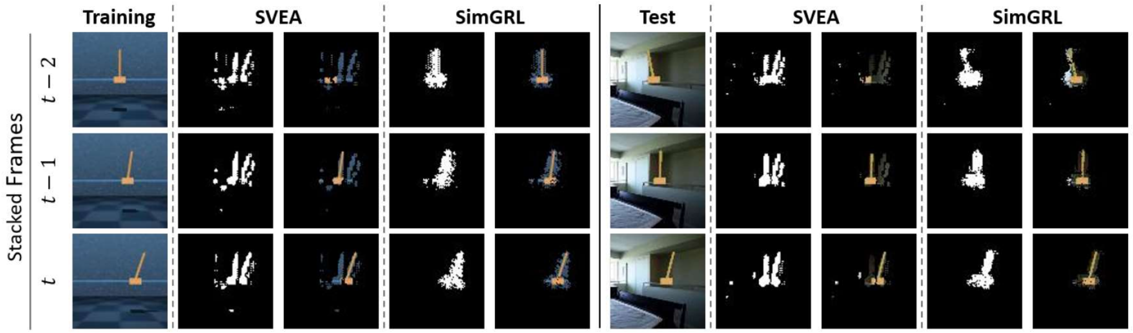


[1] "Look where you look! Saliency-guided Q-networks for generalization in visual Reinforcement Learning." *NeurIPS* (2022).
[2] "Stabilizing deep q-learning with convnets and vision transformers under data augmentation." *NeurIPS* (2021).

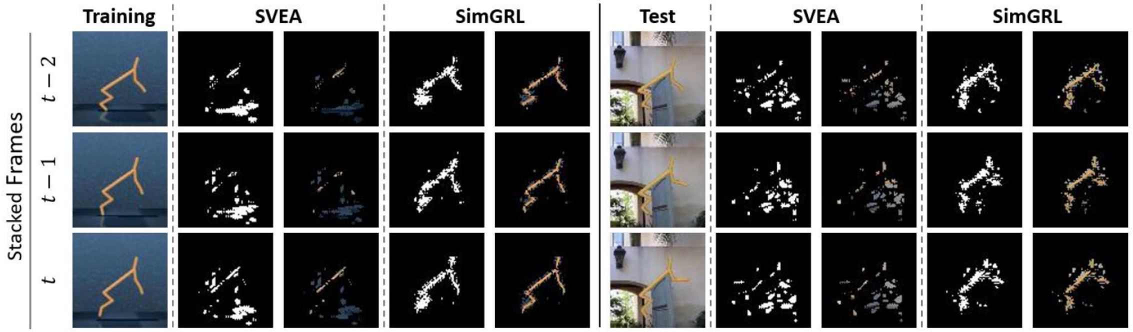
Experiments

Task-identification Capability of SimGRL

- Compared to SVEA, the proposed SimGRL accurately identifies the true salient pixels in both training and 'Video Hard' test environments of DMControl-GB.



(a) Attribution masking examples in 'Cartpole, Swingup'

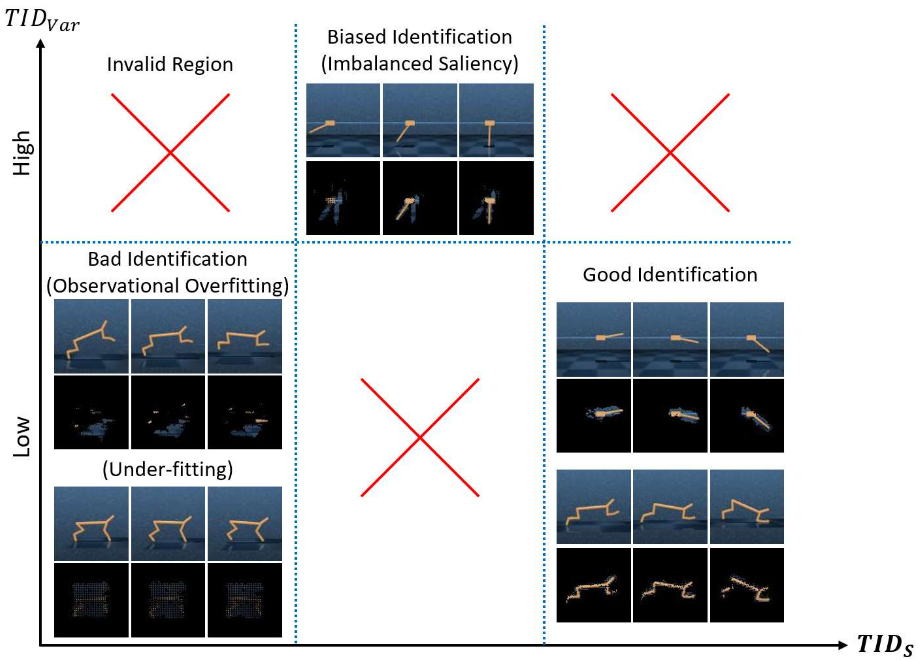


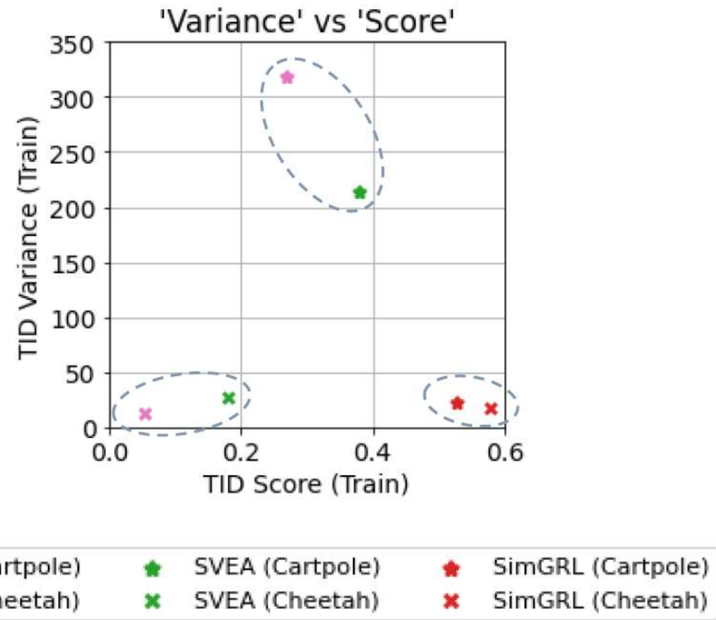
(b) Attribution masking examples in 'Cheetah, Run'

Experiments

Analysis with TID Metrics

- SimGRL shows relatively *high TID scores and low TID variances regardless of tasks*, implying the mitigation of both problems.



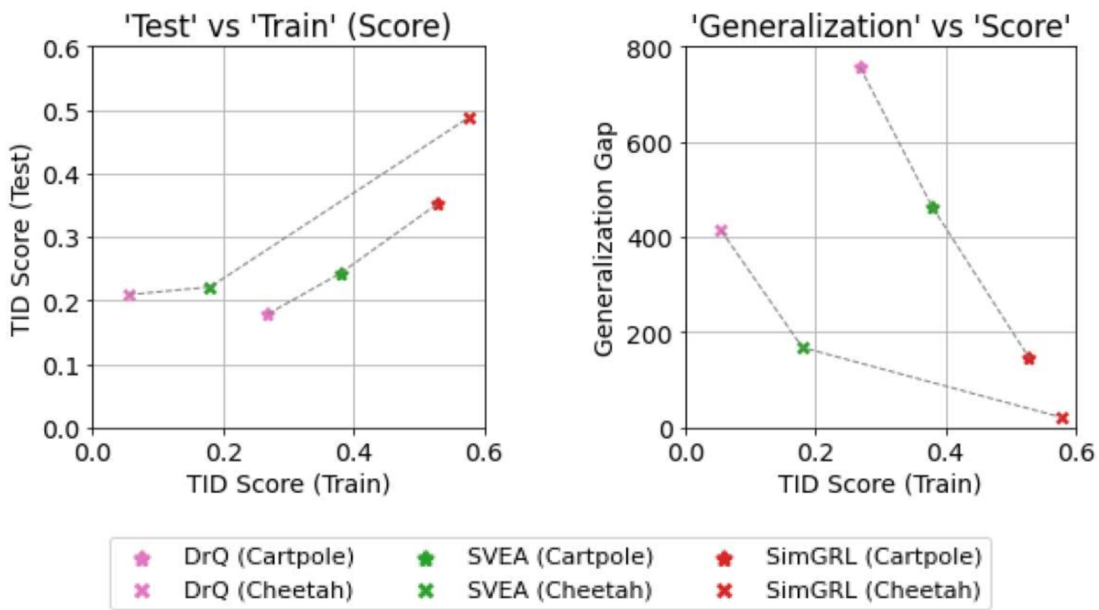


Experiments

Analysis with TID Metrics

- Good task identification in training environments can lead to :

1) Good task identification also in test environments.
2) Good generalization performance, thanks to the *reduced overfitting* to training environments.



Conclusion

- By utilizing gradient-based attribution masks, we highlight the two core issues of imbalanced saliency and observational overfitting. Additionally, we propose TID metrics to measure the discrimination ability of an RL agent on task objects, providing insights into these issues.

- To address these problems, we propose architectural and data regularization methods through a modification to an encoder structure and an introduction of new data augmentation.

- We achieve state-of-the-art performances across video benchmarks of DMControl-GB, DistractingCS, and robotic manipulation tasks.

Thank You!

Poster Session :
Thu 12 Dec 4:30 p.m. PST — 7:30 p.m. PST

Wonil Song, Ph.D.
Digital Image Media Lab.
Yonsei University, Seoul, Korea
E-mail: swonil92@yonsei.ac.kr