REINFORCEMENT LEARNING CONFERENCE - RLC 2026 - POSTER #27

segdac.github.io

#27

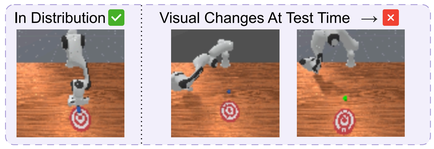
SegDAC: Visual Generalization in Reinforcement Learning via Dynamic Object Tokens

Alexandre Brown & Glen Berseth
Mila - Québec AI Institute, Université de Montréal, REAL

The problem

Image-based policies break under visual shift.

- Small pixel changes greatly reduce performance
- Even when the task structure is the same

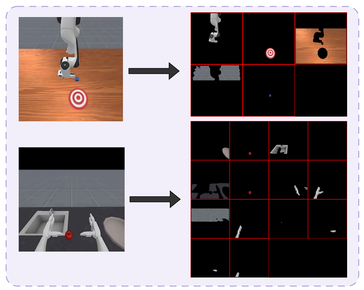


Key Ideas

Reason over objects, not pixels.

- Represent state as objects
- Train policy on object-based representation
- How ?

Dynamic: The object set is not fixed: count, identity, and size change every timestep.



Contributions

1. Efficient Policy training with support for a variable number of objects that results in increased robustness.
2. A method to construct contextual per-object tokens from frozen pretrained vision models, with segment positional encoding.
3. **+88%** improvement over prior visual-generalization methods on the hard setting, while matching DrQ-v2 sample efficiency.
4. A **visual-generalization benchmark**: 8 ManiSkill3 tasks, 12 perturbation types, 3 difficulty levels, organized by a scene-entity taxonomy.

Simple To Train

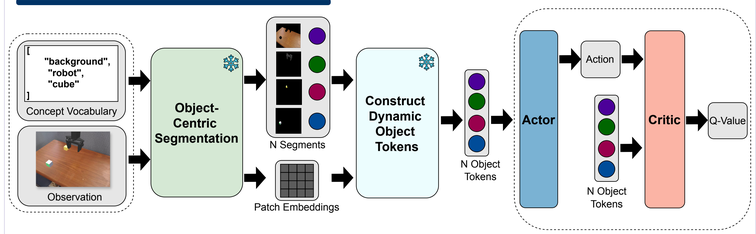
SegDAC only needs the RL reward. Plain SAC loss, drop-in for any model-free algorithm.

- ✅ **No hand-crafted data augmentation**
- ✅ **No auxiliary losses**
- ✅ **No image reconstruction**
- ✅ **No ground-truth masks**

$$J_Q(\theta) = \mathbb{E}_{(\mathbf{s}_t,\mathbf{a}_t)\sim\mathcal{D}}\left[\frac{1}{2}\left(Q_\theta(\mathbf{s}_t,\mathbf{a}_t) - \left(r(\mathbf{s}_t,\mathbf{a}_t) + \gamma\,\mathbb{E}_{\mathbf{s}_{t+1}\sim p}\left[V_{\bar{\theta}}(\mathbf{s}_{t+1})\right]\right)\right)^2\right]$$

$$J_\pi(\phi) = \mathbb{E}_{\mathbf{s}_t\sim\mathcal{D},\epsilon_t\sim\mathcal{N}}\left[\alpha\log\pi_\phi(f_\phi(\epsilon_t;\mathbf{s}_t)|\mathbf{s}_t) - Q_\theta(\mathbf{s}_t, f_\phi(\epsilon_t;\mathbf{s}_t))\right]$$

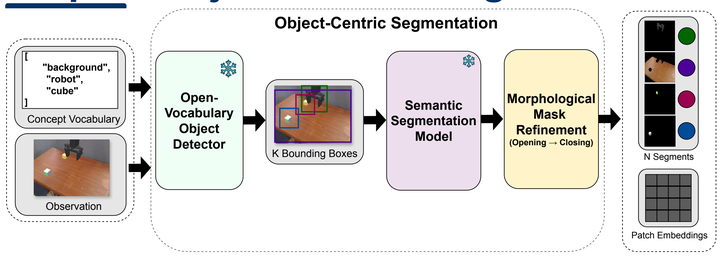
SegDAC Pipeline



SegDAC : **Seg**mentation-**D**riven **A**ctor-**C**ritic for visual RL.

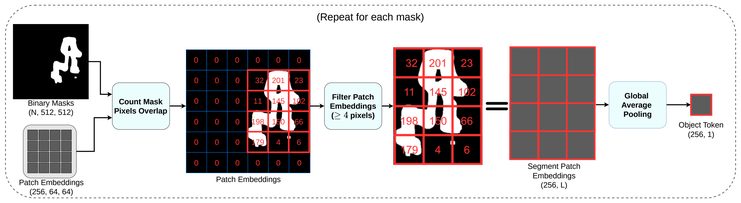
- General idea: segment the image into (sub)objects, extract object tokens, then train on the ever-changing object tokens set using model-free online RL.

Step 1 : Object-Centric Segmentation



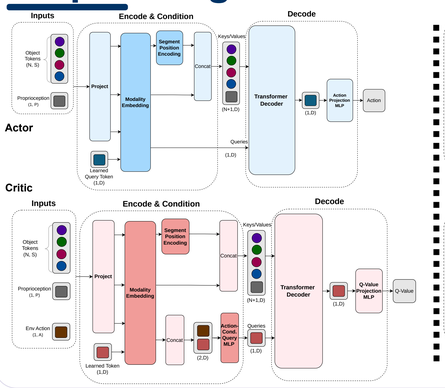
- We extract image segments efficiently using fast pretrained vision models.

Step 2 : Per-object Token Extraction



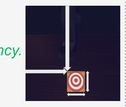
- We extract object tokens by pooling frozen patch embeddings that overlap with segment masks.
- Key design: reuse existing patch embeddings so each object token retains global context.

Step 3 : Segment-aware Actor-Critic RL



Segment Positional Encoding

Helps with sample efficiency.

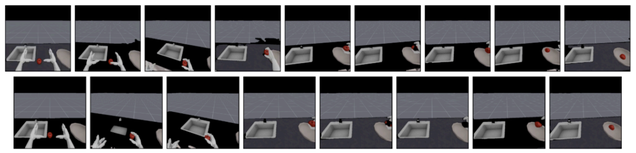


Transformer Decoders & Learned Query Fusion

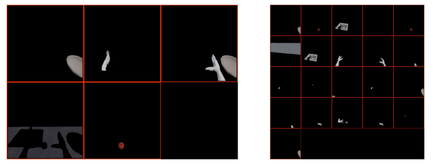
Helps with training stability.

Dynamic Object Tokens Training

- Unlike prior object-centric RL, we train in latent space on a variable-length token set stored in the replay buffer, with no fixed object count, ground-truth masks, or image reconstruction.
- Varying token count and identity acts as built-in domain randomization, keeping SegDAC's policy robust even when detection drops a task-critical object.

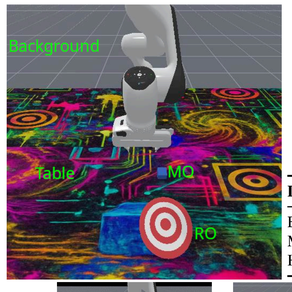


Different Timesteps Can Produce Different Numbers of (sub)-Objects



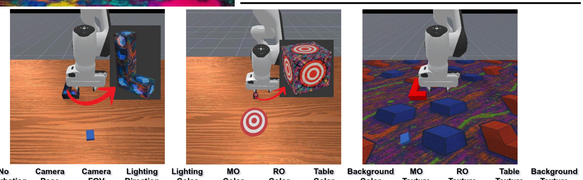
A New Visual Generalization Benchmark

8	12	3
tasks	perturbations	difficulties

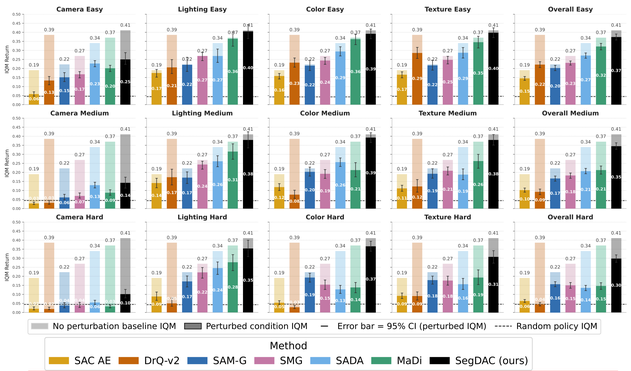


- Built on ManiSkill3: manipulation under visual generalization.
- Systematic perturbations including semantic conflicts.

Difficulty	Visual change	Semantic perturbation
Easy	✓ Low	✗ None
Medium	✓ Moderate	✗ None
Hard	✓ High	✓ Present



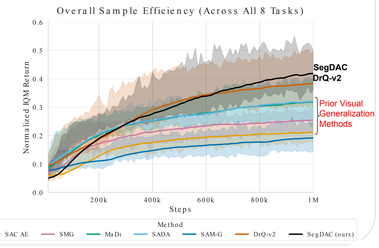
Generalization Results



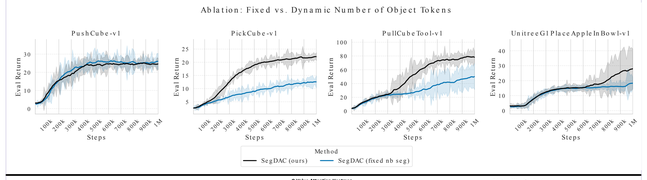
88% higher IQM return than strong visual baselines on the hardest perturbations. The gap widens as perturbations get harder.

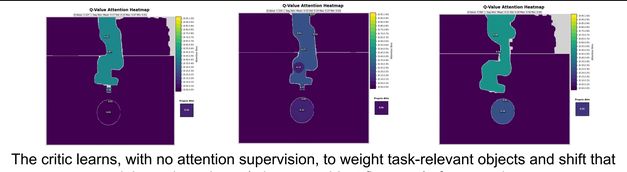
Sample Efficiency Results

Matches DrQ-v2 sample efficiency and stays stable across all 8 tasks.



Analysis & Ablations





The critic learns, with no attention supervision, to weight task-relevant objects and shift that weight as the sub-goal changes: object first, goal after grasping.

Main Takeaways

- Visual generalization in RL has largely relied on **hand crafted data augmentation**. We show strong **generalization is achievable without it**.
- **Reasoning over objects** rather than pixels is a **favorable inductive bias** for both robust visual **generalization** and efficient learning.
- **Training on a dynamic set of object tokens**, changing in count and identity every step, is stable under standard SAC and promotes **robustness to state variation**.