[Paper](#) [GitHub](#)

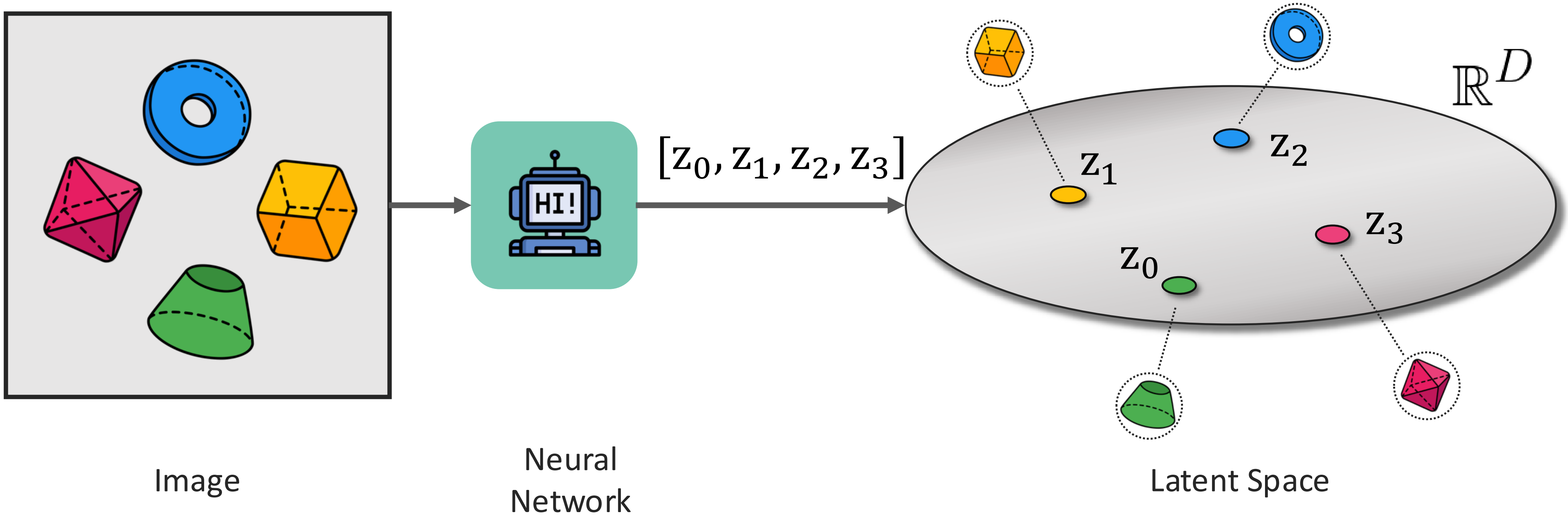
# Factor-Wise Homogeneity of Slot-Attention for Continual Object-Centric Learning

ICML 2026

Ilmin Kang , Hoyong Kim, Seungju Bang, Minwoo Kang, Kangil Kim*

# Object-Centric Learning (OCL)

- OCL aims to learn a set of disentangled representations, each corresponding to an individual object without the need for explicit supervision

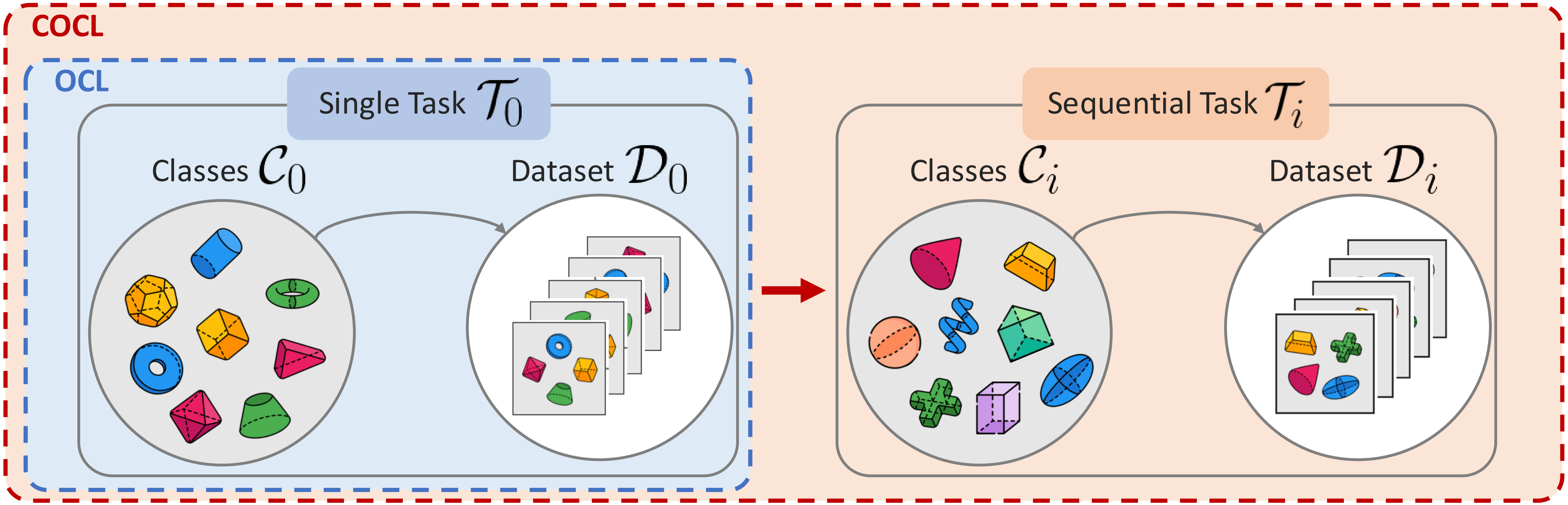


[Problem]

# OCL Beyond Single Task Learning

- OCL typically operate under the assumption of a static, single-task environment

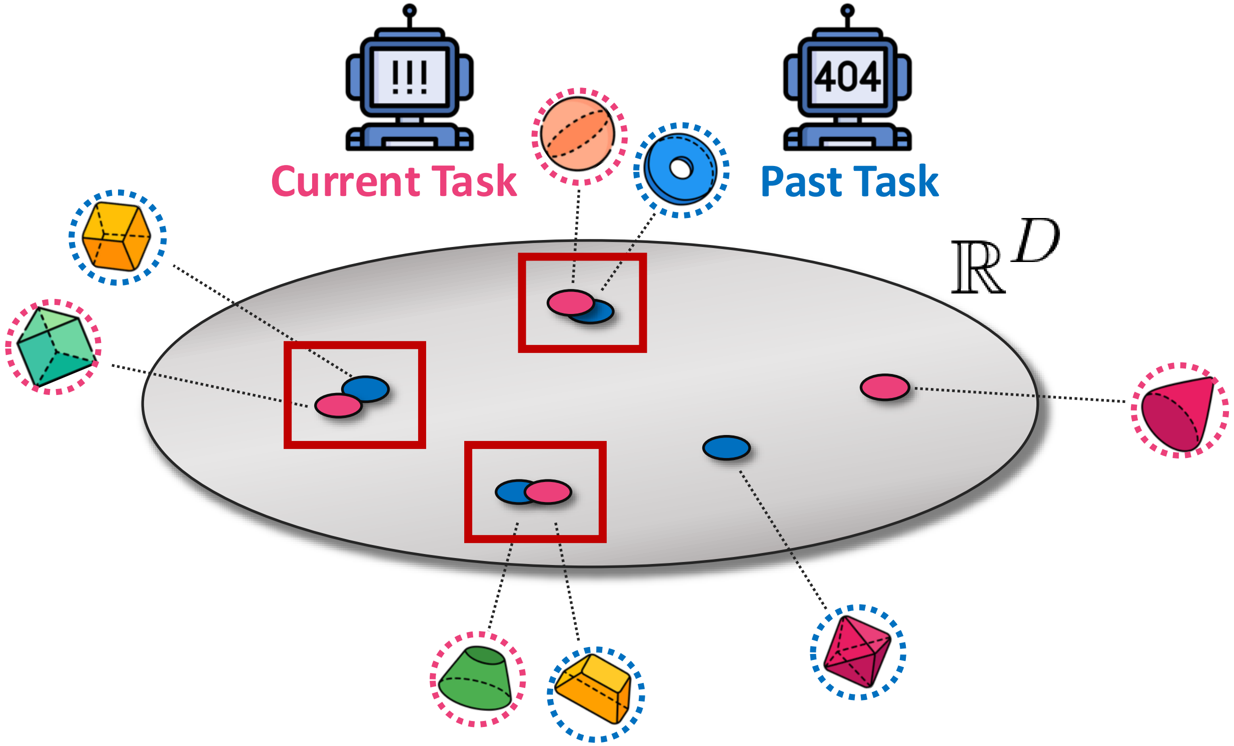
➔ We propose novel Continual Object-Centric Learning (COCL)



# Object-wise Forgetting in COCL

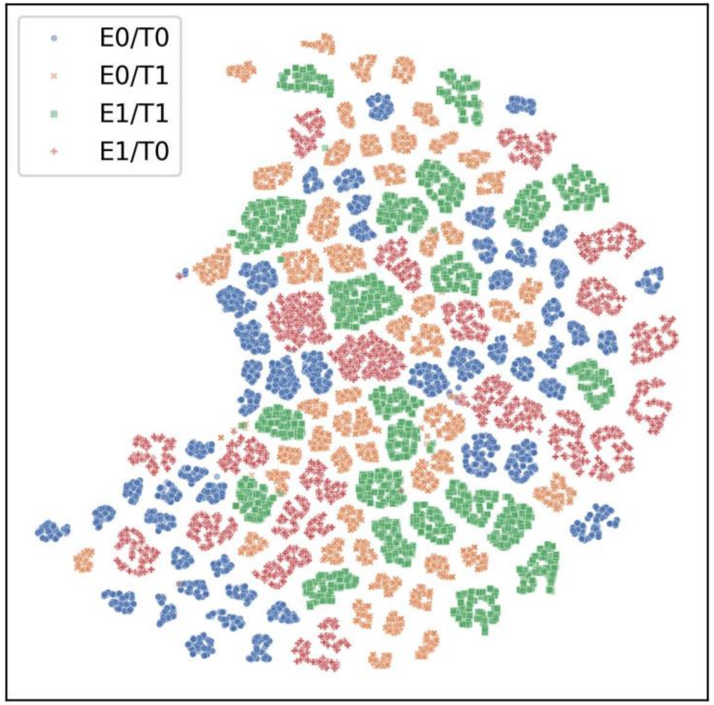
- Object-wise Catastrophic Forgetting due to inter-task overlapping

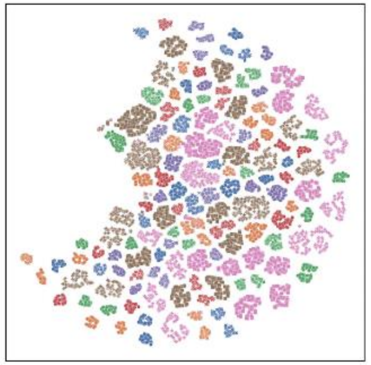
- As the model updates with new information, the latent representations of novel tasks tend to interfere with the regions occupied by previous tasks



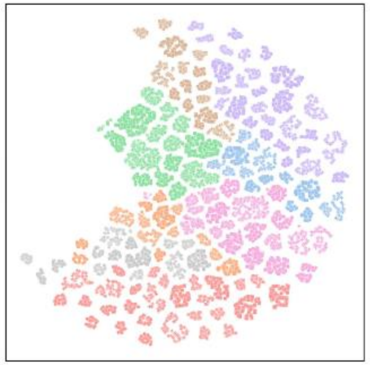
# Factor-Wise Homogeneity Property of Slot Attention

- We investigate Slot Attention and reveals its property: Factor-Wise Homogeneity
    1. Organizes latent representations into small and separated regions, each of which preserves identical factor states
    2. This emerges across sequential tasks with novel factors

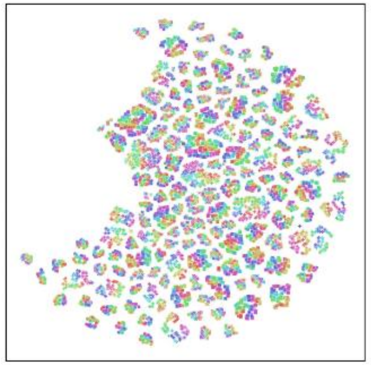


Shape



Position



Color

➔ Causing inter-task separation

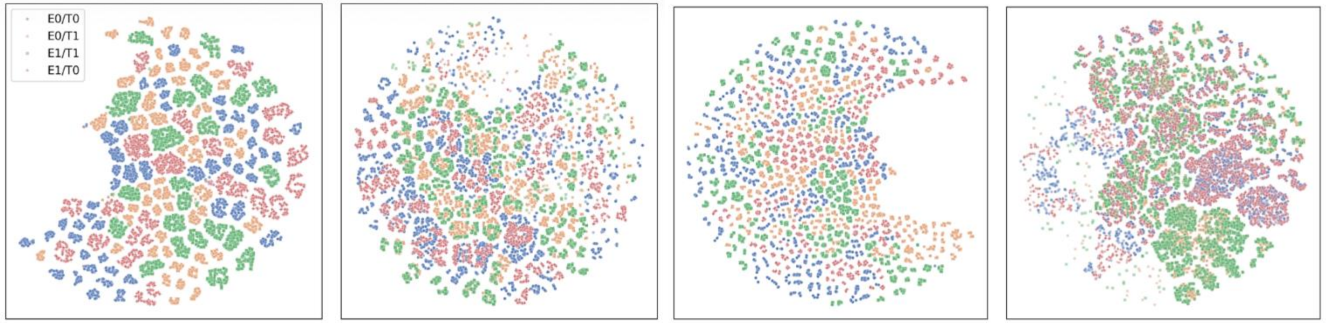
➔ Serving as a key mechanism to prevent catastrophic forgetting from inter-task overlapping

Locatello et al., 2020

[Findings]

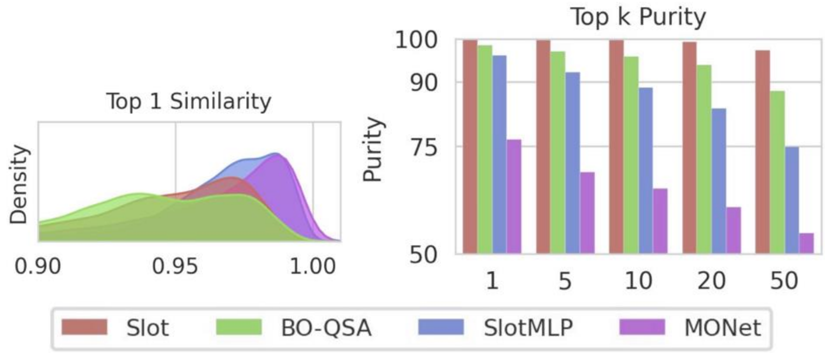
# Benefits of Factor-Wise Homogeneity to COCL

- Unique property of Slot Attention




(a) Slot Attention (b) SlotMLP (c) BO-QSA (d) MONet

***Qualitative evaluation of inter-task separation***




***Quantitative evaluation of inter-task separation***

- Mitigates interference between tasks
  - Preserves task boundaries and discriminative semantic features

| Model | Task Prediction | | Semantic Prediction | |
|---|---|---|---|---|
| | Acc. | F1 | Acc. | F1 |
| SlotMLP | 77.08 | 76.95 | 74.75 | 62.15 |
| MONet | 76.38 | 75.47 | 61.39 | 42.70 |
| BO-QSA$^*$ | 94.93 | 94.71 | 92.72 | 91.39 |
| SA$^{\dagger *}$ | **96.82** | **96.60** | **99.82** | **99.83** |

***Linear probing on latent representation***

**** Denotes models exhibiting factor-wise homogeneity***

# Bottleneck in Exploiting Inter-Task Separation

- Decoder acts as a bottleneck in exploiting inter-task separation
- Decoder inevitably overfits to the visual statistics of the current task

➔ Propose minimal yet effective strategy: Decoder-only Post Replay (DPR)

1. Partial Freezing
   → freeze the encoder and Slot Attention module, updating only the decoder
2. Post Replay
   → replay phase occurs only after the initial training on the current task

- DPR serves not as a trivial solution, but as a mechanism relying on factor-wise homogeneity
- And demonstrates the significance of this property

| Model | E0 / T0 | E0 / T1 | |
|---|---|---|---|
| | | w/o DPR | w/ DPR |
| SlotMLP | $75.83_{\pm.13}$ | $32.85_{\pm.07}$ | $45.75_{\pm.10}$(+13.) |
| MONet | $84.63_{\pm.11}$ | $43.51_{\pm.06}$ | $44.41_{\pm.04}$(+1.) |
| BO-QSA* | $99.98_{\pm.00}$ | $50.46_{\pm.08}$ | **96.54**$_{\pm.02}$(+48.) |
| SA†* | $99.86_{\pm.00}$ | $41.39_{\pm.03}$ | **98.81**$_{\pm.00}$(+57.) |

*Results of object discovery task (FG-ARI) * Denotes models exhibiting factor-wise homogeneity*

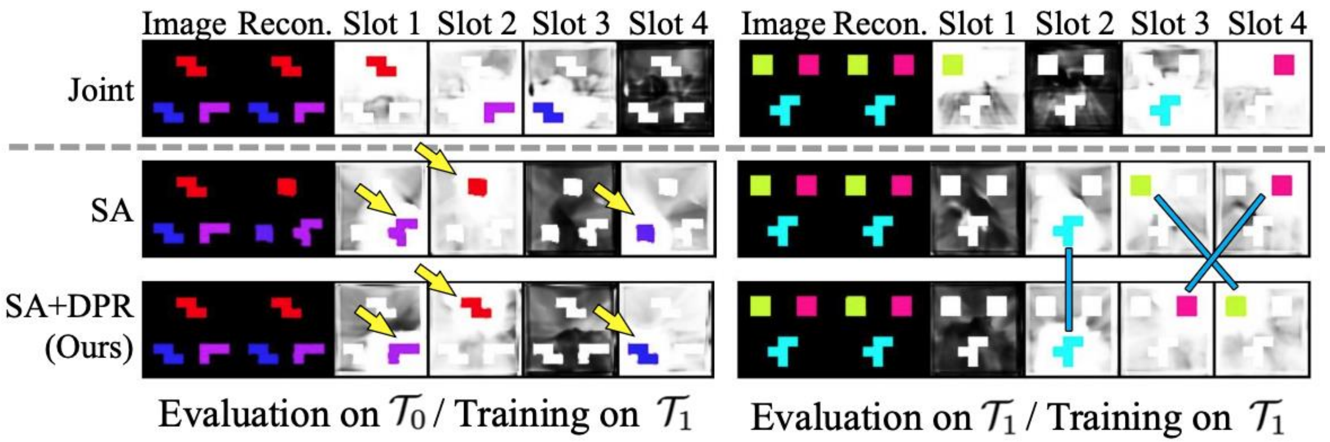
# Performance of Factor-wise Homogeneity with DPR

| Model | E0 / T0 | E0 / T1 | E1 / T1 |
|---|---|---|---|
| *Dataset: C-Tetrominoes SST* | | | |
| SA†(Joint) | $99.89_{\pm.00}$ | | $99.85_{\pm.00}$ |
| SA† | $99.86_{\pm.00}$ | $41.71_{\pm.03}$ | $99.79_{\pm.00}$ |
| SA†+ DPR | $99.88_{\pm.00}$ | **98.81**$_{\pm.00}$(+57.) | $99.70_{\pm.00}$ |
| *Dataset: C-CLEVR SST* | | | |
| SA†(Joint) | $96.24_{\pm.01}$ | | $96.66_{\pm.01}$ |
| SA† | $96.01_{\pm.01}$ | $63.83_{\pm.03}$ | $96.21_{\pm.01}$ |
| SA†+ DPR | $96.03_{\pm.01}$ | **92.00**$_{\pm.01}$(+28.) | $95.36_{\pm.01}$ |

***COCL benchmarks (shape)***

| Model | E0 / T0 | E0 / T1 | E1 / T1 |
|---|---|---|---|
| *Dataset: C-Tetrominoes SST-(color)* | | | |
| SA† | $99.88_{\pm.00}$ | $70.45_{\pm.06}$ | $96.85_{\pm.05}$ |
| SA†+ DPR | $99.95_{\pm.01}$ | **96.84**$_{\pm.00}$(+26.) | $99.42_{\pm.00}$ |
| *Dataset: C-Tetrominoes SST-(shape+color)* | | | |
| SA† | $99.94_{\pm.00}$ | $48.03_{\pm.01}$ | $99.90_{\pm.00}$ |
| SA†+ DPR | $99.95_{\pm.00}$ | **94.19**$_{\pm.02}$(+46.) | $99.51_{\pm.00}$ |

***Robustness to different semantic factors***



# Performance of Factor-wise Homogeneity with DPR

| Dataset | Model | E0 / T0 | | E0 / T3 | |
|---|---|---|---|---|---|
| | | FG-ARI | $\text{mBO}^c/\text{mBO}^i$ | FG-ARI | $\text{mBO}^c/\text{mBO}^i$ |
| COCO | DINOSAUR | $23.01_{\pm 0.8}$ | $45.87_{\pm 0.4}$ / $35.43_{\pm 0.4}$ | $21.25_{\pm 0.6}$ | $42.14_{\pm 0.6}$ / $32.76_{\pm 0.6}$ |
| | DPR†† | | | **23.03**$_{\pm 0.8}$ | **43.47**$_{\pm 0.5}$ / **33.80**$_{\pm 0.6}$ |
| PASCAL | DINOSAUR | $18.01_{\pm 0.6}$ | $55.06_{\pm 0.8}$ / $50.11_{\pm 0.4}$ | $14.63_{\pm 0.7}$ | $51.47_{\pm 0.8}$ / $47.26_{\pm 0.4}$ |
| | DPR†† | | | **16.69**$_{\pm 0.6}$ | **52.92**$_{\pm 0.7}$ / **48.56**$_{\pm 0.5}$ |

***Scalability to complex real-world dataset***

| Method | Memory | Replay | FG-ARI | |
|---|---|---|---|---|
| | | | E0 / T0 | E3 / T1 |
| DINOSAUR | - | - | $23.54_{\pm .03}$ | $22.27_{\pm .02}$ |
| DDGR | DDGR | RS | $23.54_{\pm .03}$ | $24.08_{\pm .04}$(1.80) |
| DPR | Random | DPR | | $23.63_{\pm .01}$(1.35) |
| DPR† | DDGR | DPR | | $24.03_{\pm .01}$(1.76) |
| DPR‡ | DDGR | RS+DPR | | **24.25**$_{\pm .05}$(1.98) |

***Comparable performances with diffusion generative replays***

| Model | FWH | DPR | C-Tetrominoes | C-CLEVR |
|---|---|---|---|---|
| SA† | O | X | $54.49_{\pm .01}$ | $58.68_{\pm .03}$ |
| SA† | O | O | **98.70**$_{\pm .04}$(+44.2) | **90.72**$_{\pm .05}$(+32.0) |
| Monet | X | X | $53.57_{\pm .05}$ | $58.76_{\pm .09}$ |
| Monet | X | O | $55.00_{\pm .06}$(+1.4) | $61.36_{\pm .04}$(+2.6) |

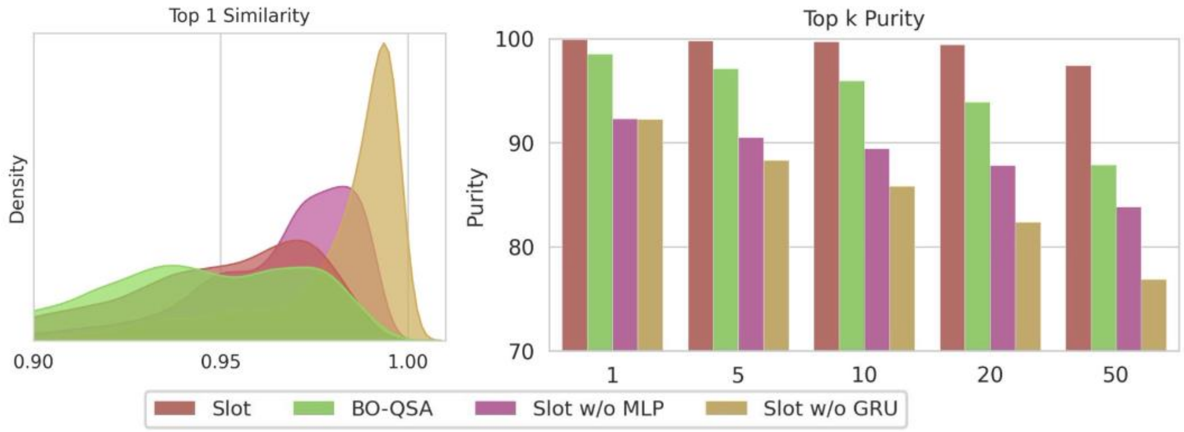
| Model | FHW | Task Prediction | | Semantic Prediction | |
|---|---|---|---|---|---|
| | | Acc. | F1 | Acc. | F1 |
| MONet | X | 72.37 | 71.87 | 60.37 | 60.97 |
| SA† | O | **92.36** | **90.14** | **88.72** | **87.95** |

***Scalability to open world scenarios***

[Analysis]

# What Drives Factor-wise Homogeneity?

- Module Influence on Slot Organization
  - 1. Random initialization, 2. Attention, 3. Gated Recurrent Unit (GRU), 4. MLP




*Ablation study w.r.t. inter-task separation*

| | Attention | GRU | MLP |
|---|---|---|---|
| Norm | 1131.06 | 1380.01 | 761.01 |
| Rank ($D = 64$) | 23.91 | 31.17 | 34.21 |

*Analysis of the influence of GRU to slot organization using Singular Value Decomposition on the Jacobian matrix*

➔ 1. Slot Attention's GRU serves as the core mechanism for the organization of its representations

- Object-wise Consistency of GRU
  - Slow mode of Recurrent Neural Networks
    - Gates generate marginally stable directions
    - Slowly changes hidden-states
    - Provides an inherent mechanism for maintaining global information over time
  - Spearman's rank correlation of this Slow mode

| (E$i$)/(T$j$) | Identical | Non-Identical |
|---|---|---|
| E0 / T0 | 0.9921 | 0.7931 |
| E1 / T1 | 0.9920 | 0.7958 |
| E0 / T1 | 0.9935 | 0.8054 |
| E0 / T1 VS E1 / T1 | - | 0.7929 |

*Factor-wise consistency of GRU's slow mode ranks.*

➔ 2. GRU structurally aligns identical semantic factors into highly stable dimensions while separating differing factors from sequential tasks

Krishnamurthy et al., 2022

# Takeaways

- Factor-wise homogeneity property of Slot Attention
  - Organizes the slot representations into well-separated regions that preserve identical semantics across tasks
  - This establishes an intrinsic inter-task separation, benefiting inter-task interference in continual learning
- However, decoder shows bottleneck in exploiting this separation
  - We propose minimal strategy Decoder-only Post-Replay which can simply address this limitation
- GRU are the essential for both factor-wise homogeneity and inter-task separation

# Thank You

For details of our proposal, check out our Paper & GitHub

See you at Seoul!

Paper GitHub