Multi-Part Object Representations via Graph Structures and Co-Part Discovery

A. Foo, W. Hsu, and M. L. Lee
School of Computing, National University of Singapore

Motivation

(a) Information-bottleneck approaches struggle to discriminate object boundaries.

(b) Methods that cluster learned features struggle to compose part clusters into object wholes.

(c) Proposed Explicit Compositional Network (ECO-Net) robustly composes parts into object wholes.

ECO-Net: Overview

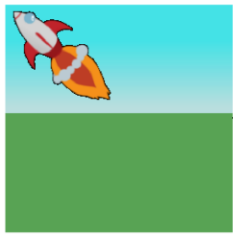


Image 1

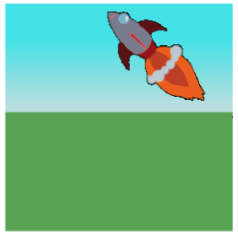
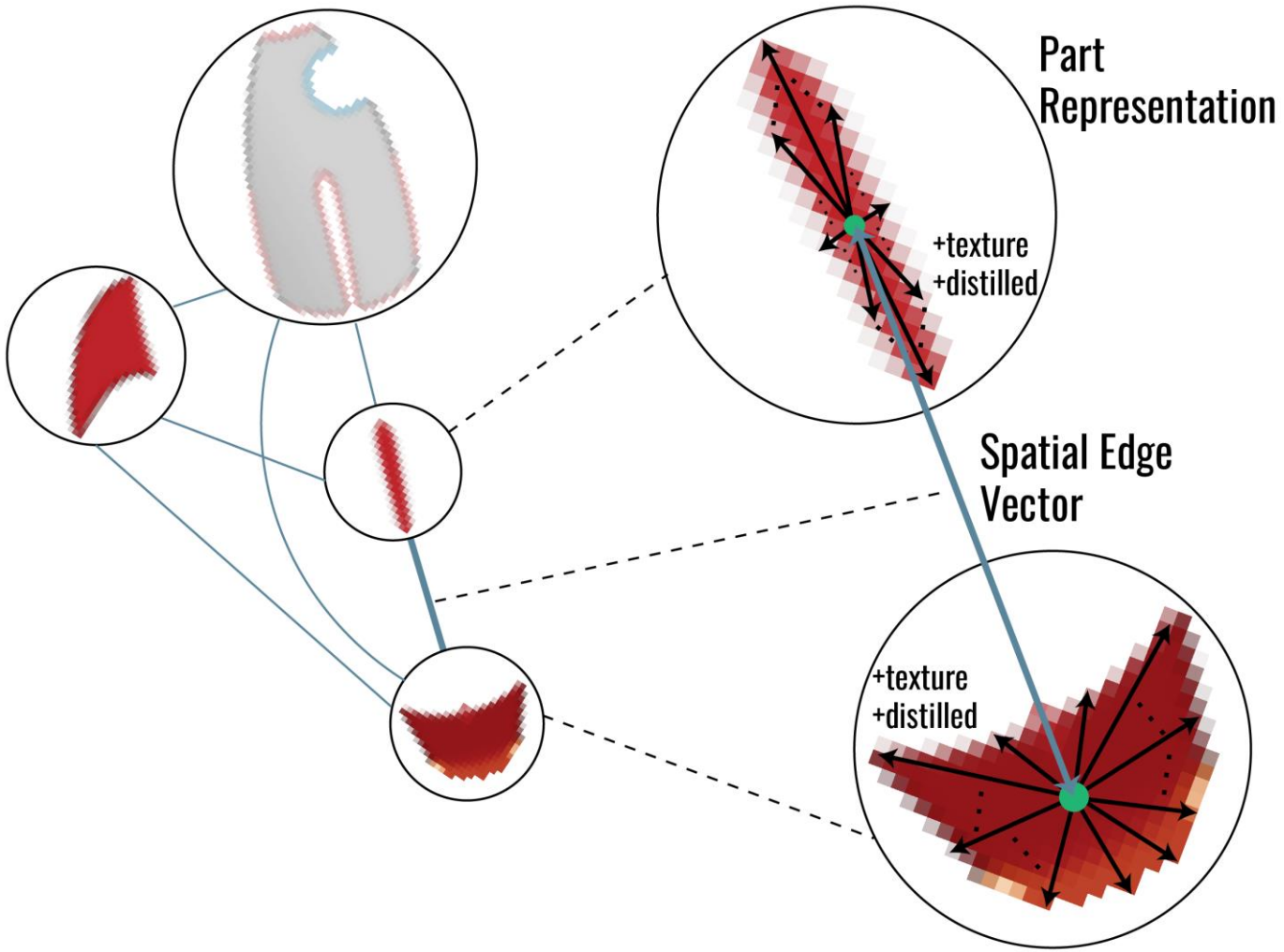


Image 2

ECO-Net: Graph Representation

First construct Graph of Parts via a fine-grained segmentation algorithm (e.g. Felzenswalb's)

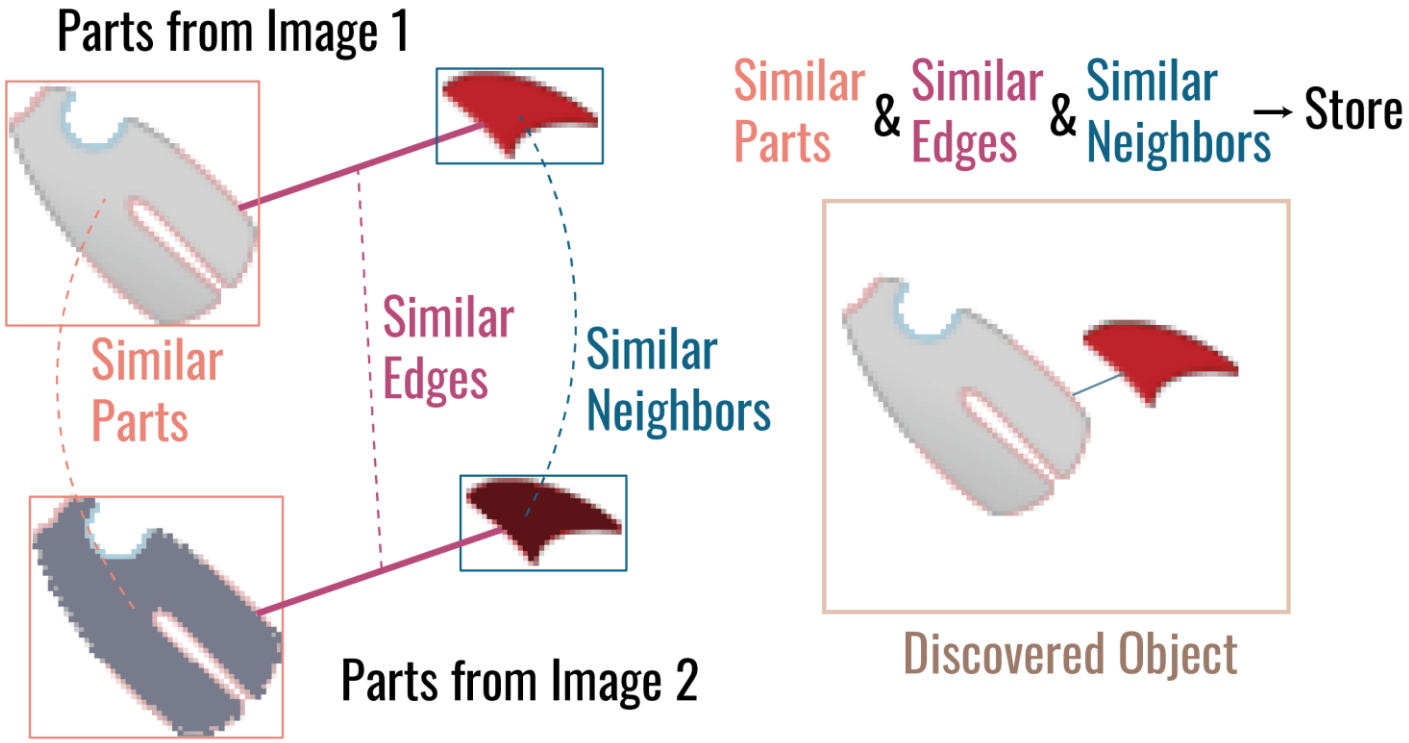
- Each node summarizes the position-invariant shape (using vectors from its centroid to the sampled boundary pixels), texture and distilled features of each part.
- Each edge depicts the spatial relationships between adjacent parts.



ECO-Net: Co-Part Object Discovery

Based on the intuition that neighboring parts that recurringly appear together across samples should belong to the same object [Brady and Kersten 2003; Kruger et al. 2012]

Repeated occurrence of parts with similar spatial relationships are grouped into objects.

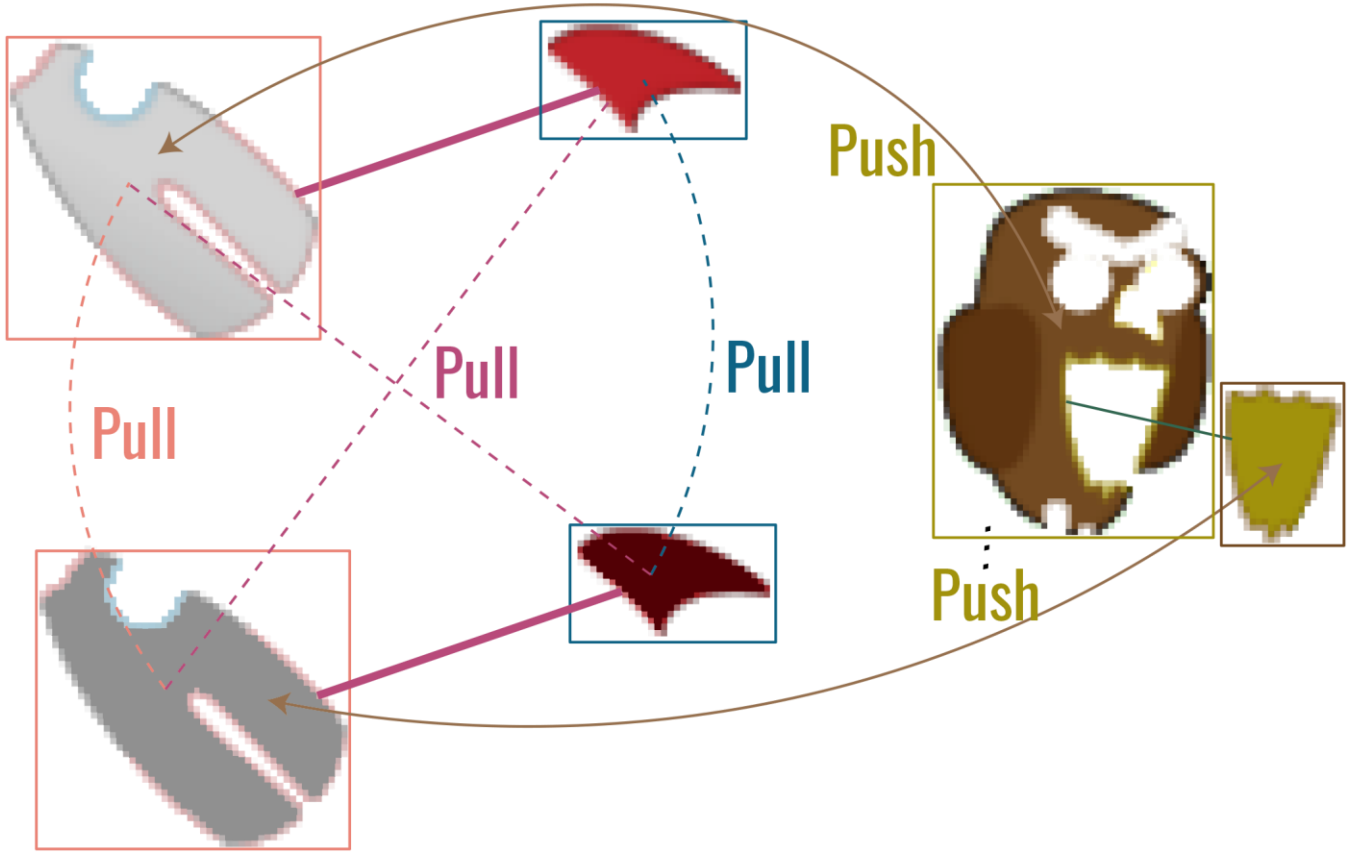


ECO-Net: Part-Centric Regularization

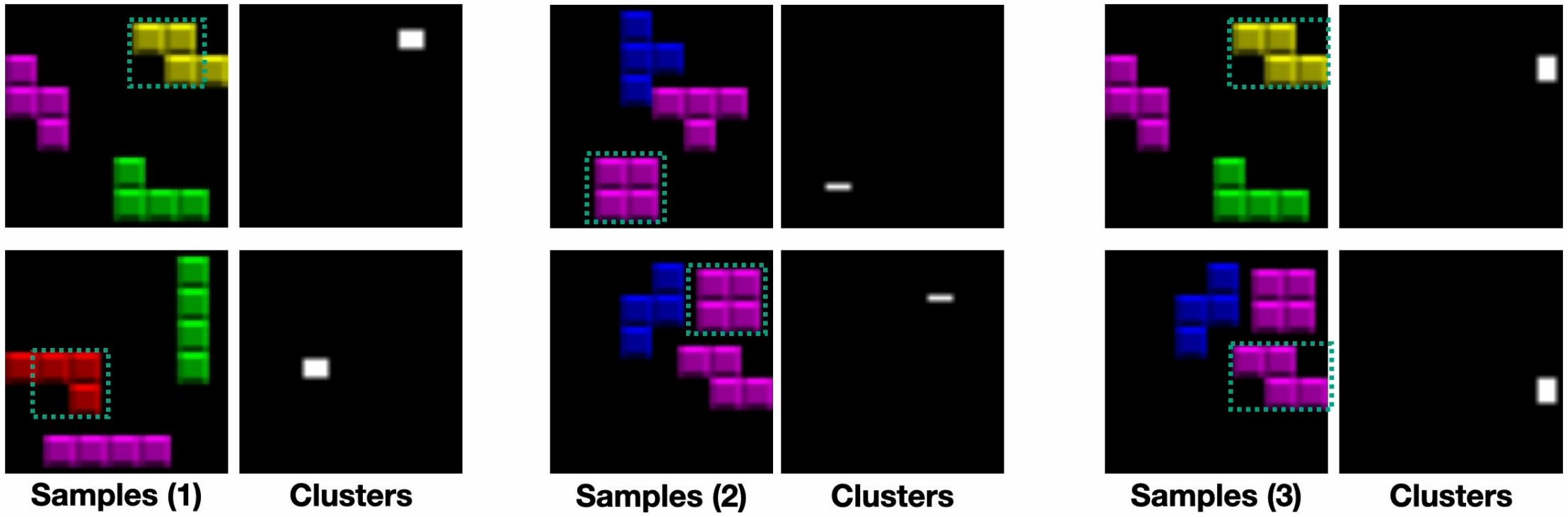
Finally, to improve consistency of object discovery across images, we refine part representations by:

- Maximizing similarity between parts assigned to the same object across instances.

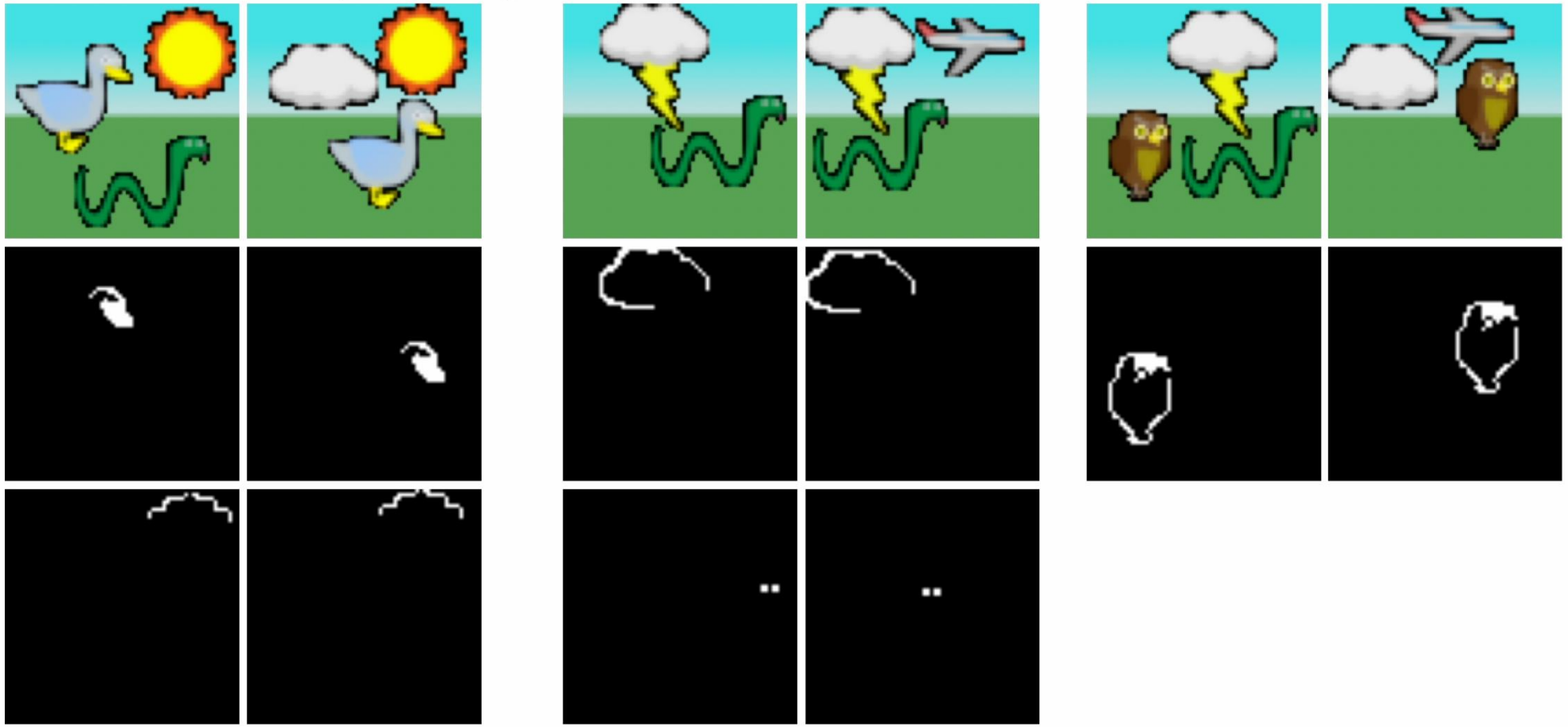
- Minimize similarity between parts belonging to different objects.



ECO-Net: Co-Part Object Discovery

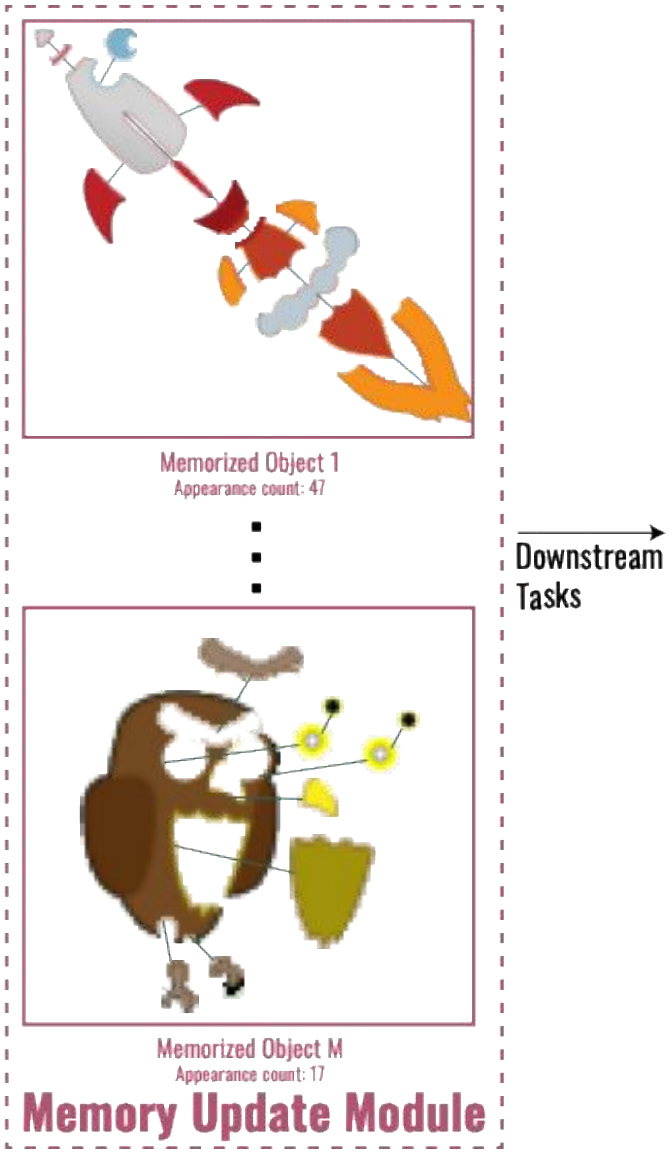


ECO-Net: Co-Part Object Discovery



ECO-Net: Memory Update

- Obtain a position-invariant object representation
- If this object already exists in memory, we increase the appearance count. Otherwise, we add this new object to memory.
- Once all training samples have been processed, we sort the memory of objects by the number of appearances to determine the priority of memorized objects.
- The memory can then be used for downstream object discovery



Experiment Setup: Datasets

Dataset	Type	Ground Truth	Image Size	# Samples	Min Visible
Tetrominoes	Simulated	Pixel Mask	35×35	63K	100%
AbsScene	Simulated	Pixel Mask	64×64	63K	100%
GSO	Realistic	Pixel Mask	128×128	63K	100%
SKU-110K	Realistic	Bounding Box	128×128	11K	Varied
PASCAL VOC	Real-World	Pixel Mask	128×128	12K	Varied
MS COCO	Real-World	Pixel Mask	128×128	12K	Varied
AbsScene-O	Occluded Objects	Pixel Mask	64×64	63K	25%
GSO-O	Occluded Objects	Pixel Mask	128×128	63K	25%
AbsScene-C	Out-of-Distribution	Pixel Mask	64×64	1K	100%

Experiments on Quality of Discovered Objects

(a) Simulated datasets

Method	Tetrominoes			AbsScene		
	ARI	mDice	mIoU	ARI	mDice	mIoU
SLIC	51.7±17.2	65.2±14.8	52.5±16.2	26.2±10.4	53.9±9.4	38.3±9.3
Ncut	66.4±19.7	62.8±17.8	56.5±19.1	23.7±17.8	25.7±8.9	15.7±6.9
Felzenszwalb	91.9±18.2	96.6±7.8	95.0±11.2	76.8±30.3	85.8±14.1	78.2±17.2
Slot Attention	99.7±1.5	41.4±1.5	26.6±1.5	94.8±11.8	74.6±10.8	63.9±13.4
GENESIS-V2	97.2±5.4	47.0±3.5	30.9±3.0	46.7±24.1	65.4±15.8	54.3±18.0
BO-QSA	99.1±3.7	40.8±1.4	25.8±1.2	93.4±13.8	83.1±9.2	74.0±12.4
DINOSAUR	92.5±9.9	40.0±2.0	25.3±1.8	95.0±8.2	51.9±6.2	35.5±5.7
MaskCut	93.4±15.3	80.8±5.6	68.5±6.6	90.5±18.9	88.3±10.7	81.5±12.8
OC-Net	92.4±15.4	96.7±6.5	94.9±9.7	78.5±24.7	86.4±11.7	78.7±14.4
ECO-Net	**100.0±0.0**	**100.0±0.0**	**100.0±0.0**	**97.7±2.4**	**99.4±1.0**	**98.8±1.5**

(b) Realistic datasets

Method	GSO			SKU-110K		
	ARI	mDice	mIoU	ARI	mDice	mIoU
SLIC	18.5±16.8	41.1±14.0	27.4±12.8	2.7±2.0	1.7±0.7	0.8±0.3
Ncut	29.8±2.04	33.4±14.0	23.5±12.4	5.2±2.6	8.7±4.0	5.5±2.8
Felzenszwalb	61.2±34.3	70.3±16.9	58.8±18.1	-0.8±7.4	7.7±2.9	4.9±2.0
Slot Attention	64.3±28.8	39.5±13.2	27.0±10.8	-9.7±7.2	0.9±0.4	0.4±0.2
GENESIS-V2	90.1±8.3	84.3±4.3	73.6±5.9	6.1±2.7	6.4±2.2	3.5±1.3
BO-QSA	64.4±17.0	74.2±4.3	59.7±5.4	5.4±2.3	7.2±3.1	3.9±1.8
DINOSAUR	96.2±10.1	42.0±6.2	26.9±5.2	4.4±2.3	12.1±3.7	6.8±2.3
MaskCut	97.5±5.6	90.9±1.8	83.4±3.0	3.7±10.0	0.7±1.1	0.3±0.7
OC-Net	86.6±22.9	86.1±10.2	77.9±13.1	7.4±3.3	13.2±4.3	7.7±2.7
ECO-Net	**99.1±1.7**	**94.5±2.1**	**89.8±3.8**	**7.8±3.2**	**31.9±9.2**	**21.6±6.8**

Experiments on Quality of Discovered Objects

(c) Real-world datasets

Method	PASCAL VOC			MS COCO		
	ARI	mDice	mIoU	ARI	mDice	mIoU
SLIC	4.4±8.4	12.7±8.2	7.7±5.6	14.4±12.8	33.1±11.4	22.2±9.4
Ncut	9.2±15.4	10.6±6.6	6.7±4.8	12.9±17.1	27.2±16.2	19.3±13.8
Felzenszwalb	9.2±22.8	13.6±11.5	9.4±9.6	21.1±23.4	28.6±20.4	21.6±17.5
Slot Attention	14.8±22.4	12.6±8.3	7.7±5.6	28.5±14.2	23.9±10.2	14.5±7.1
GENESIS-V2	13.2±20.1	14.6±9.5	9.4±6.8	13.2±12.5	21.8±9.7	13.9±7.2
BO-QSA	13.0±19.5	13.7±8.9	8.6±6.2	24.0±15.3	33.3±11.6	22.5±9.4
DINOSAUR	16.9±24.0	17.0±10.5	11.4±7.7	30.6±16.9	23.0±9.3	14.2±6.6
MaskCut	17.7±28.5	19.2±15.0	15.5±13.7	37.5±23.4	43.0±18.6	33.2±16.2
OC-Net	16.6±25.7	20.7±13.9	15.1±11.6	26.5±19.3	48.1±13.6	35.5±15.1
ECO-Net	**19.0±21.4**	**27.9±14.4**	**19.9±11.7**	**38.9±18.9**	**63.4±12.7**	**49.7±13.2**

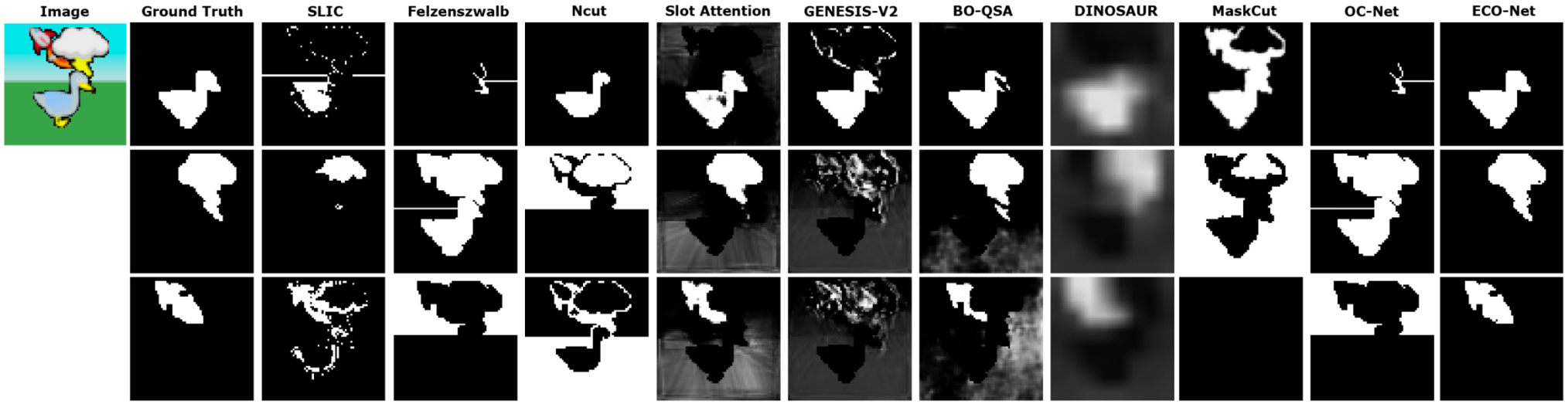
Experiments on Quality of Discovered Objects



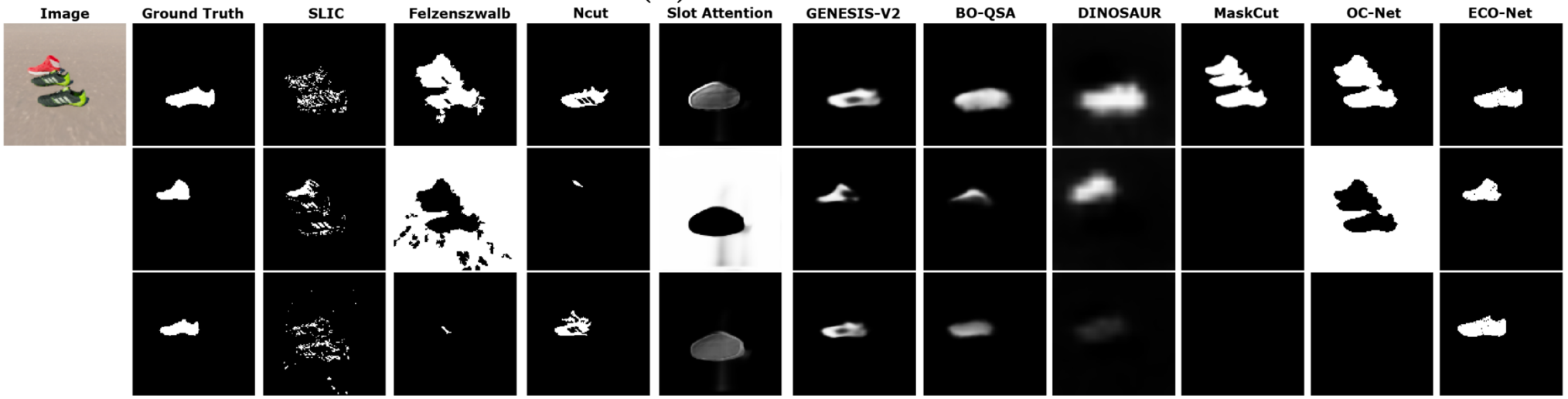
Experiments on Occlusion-Aware Perception

Method	AbsScene-O		GSO-O	
	mDice	mIoU	mDice	mIoU
SLIC	53.1±8.8	37.4±8.6	43.1±13.6	29.2±12.2
Ncut	26.8±9.4	16.6±7.4	34.0±15.4	23.3±13.7
Felzenszwalb	70.4±20.7	59.5±22.1	71.6±19.8	60.3±20.9
Slot Attention	54.7±14.4	41.6±15.0	40.5±15.7	28.1±12.4
GENESIS-V2	48.9±19.4	40.3±20.6	81.4±5.7	69.4±7.0
BO-QSA	61.0±13.7	47.1±14.9	72.2±6.3	57.3±7.6
DINOSAUR	44.5±6.4	29.1±5.3	30.0±5.6	17.8±4.1
MaskCut	68.6±21.8	59.1±23.9	85.8±3.6	76.9±4.6
OC-Net	69.9±21.0	59.2±22.5	84.5±5.3	76.2±7.0
ECO-Net	**93.1±9.6**	**90.5±12.7**	**90.3±6.7**	**83.3±9.1**

Experiments on Occlusion-Aware Perception



(a) AbsScene-O



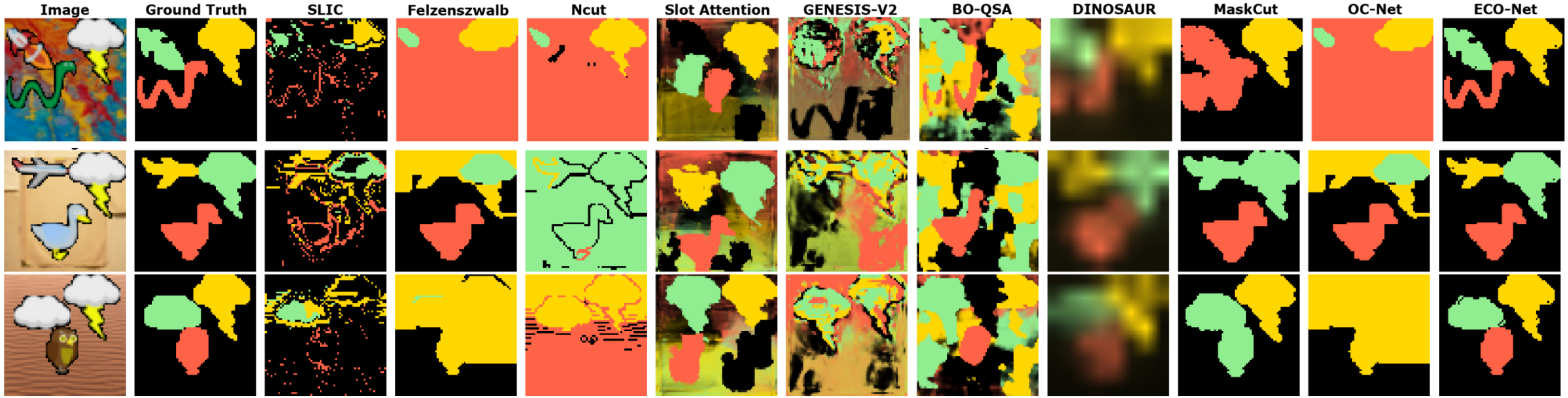
(b) GSO-O

Experiments on Out-of-Distribution Discovery

- In this set of experiments, we show that the learned object representations are robust and can be used to discover objects despite shifts in image distribution.
- We use the learned object representations from the AbsScene dataset to discover the objects in scenes with random real-world complex-texture backgrounds in AbsScene-C.

Method	ARI	mDice	mIoU
SLIC	21.0±8.6	45.8±8.9	30.8±8.0
Ncut	40.6±23.6	51.7±18.9	40.9±19.4
Felzenszwalb	60.7±34.0	68.6±25.1	60.8±27.1
Slot Attention	37.9±21.0	41.4±12.4	28.0±11.5
GENESIS-V2	23.2±14.3	32.9±8.3	20.5±6.3
BO-QSA	32.5±16.9	38.3±10.0	25.0±8.6
DINOSAUR	82.0±17.5	54.2±6.3	37.8±6.0
MaskCut	67.0±39.9	74.9±17.8	64.0±20.0
OC-Net	60.8±34.2	69.0±24.5	61.1±26.7
ECO-Net	**93.5±7.5**	**97.3±4.1**	**95.4±6.5**

Experiments on Out-of-Distribution Discovery



ECO-Net

- ECO-Net designs explicit graph-based representations and a new co-part discovery algorithm to iteratively build up objects from their parts.
- Extensive experiments demonstrate the superior quality of the objects discovered over state-of-the-art.
- Achieves robustness to occlusion and out-of-distribution settings while SOTA methods fail.

